الجمهورية الجزائرية الديمقراطية الشعبية

Ministry of Higher Education and Scientific Research

Ministry of Higher Education and Scientific Research
University of Biskra
Faculty of Exact Sciences
Department of Mathematics

وزارة التعليم العالي والبحث العلمي
جامعة بسكرة
كلية العلوم الدقيقة
قسم الرياضيات

الإحصـــــــاء الاستدلالي 1

Inferential Statistics 1

مطبوعة بيداغوجية موجهة لطلبة السنة الثانية ليسانس علم اجتماع L.M.D

Pedagogical handout for second-year undergraduate students (L.M.D) in Sociology

إعداد الدكتورة: خميسي زهية

Prepared by Dr. Khemissi Zahia

2025/2024

Table of Contents فهرس المحتويات

Content	page	المحتوى
Preface	أ-ج	المقدمة
Chapter 1 :Concept framework of inferential statistics	1	-الفصل الأول: الإطار المفاهيمي للاحصاء الاستدلالي
1.1 The Importance of Inferential Statistics in Scientific Research	2	1.1 أهمية الاحصاء الاستدلالي في البحث العلمي
1.2The Relationship of Inferential St atistics to the Social Sciences	3	1. 2 علاقة الاحصاء الاستدلالي بالعلوم الاجتماعية
1-3 Basic concepts in statistics	4	3.1المفاهيم الاساسية في الاحصاء
1-4 Statistical steps	8	4.1 الخطوات الاحصائية
1.5 Frequency tables and graphs	11	5.1الجداول التكرارية و التمثيلات البيانية
1-6-Exercies	27	6.1 تمارين
Chapter 2: Theoretical Foundations and Methodological Criteria for Hypothesis Testing	30	الفصل الثاني : الاسس النظرية و المعايير المنهجية لاختبار الفرضيات
2.1 Probability distributions in statistical inference	31	1.2 التوزيعات الاحتمالية في الاحصاء الاستدلالي
2. 2Hypothesis testing and statistical decision-making	44	2.2 اختبار الفرضيات واتخاذ ال قرارات الإحصائية
2.3Methodological Criteria Governing the Selection of Statistical Test	54	3.2المعايير المنهجية الحاكمة لاختيار الاختبار الإحصائي
2.4 Exercies	58	4.2تمارين
Chapter 3: Correlation Coefficients	59	الفصل الثالث: معاملات الارتباط

3.1Fundamentals of correl ation coefficie	**60**	1.3 اساسيات معاملات الارتباط
3.2 Correlation coefficients for quantitative variables	**62**	2.3معاملات الارتباط للمتغيرات الكمية
3.2.1 Pearson correlation coefficient	**62**	1.2.3 معامل الارتباط بيرسون
3.2.2 Partial correlation coefficient	**68**	2.2.3معامل الارتباط الجزئي
3. 3 Correlation coefficients for qualitative variables	**72**	3.3 معاملات الارتباط للمتغيرات النوعية
3.3.1Rank correlation coefficient (Spearman)	**73**	1.3.3 معامل لرتباط الرتب (سبيرمان)
3.3.2 Kandel correlation coefficient	**77**	2.3.3 معامل الارتباط كاندل
3.4 Exercies	**83**	4.3تمارين
Chapter 4 : Applications Test of the Chi-Two (Khi-2)	**87**	**الفصل الرابع: تطبيقات اختبار كاي مربع(Khi-deux / Chi²)**
4.1 The Chi-Two Test for Good Fit	**88**	1.4اختبار كاي-2 لحسن المطابقة
4.2 The Chi-Two Test for Independence	**94**	2.4 اختبار كاي-2 للاستقلالية
4.3 Exercies	**100**	3.4تمارين
Chapter 5: Linear regression models	**104**	**الفصل الخامس: نماذج الانحدار الخطي**

5.1 Fundamentals of Linear Regression	**105**	1.5اساسيات الانحدار الخطي
5.2 Simple Linear Regression Coefficient	**108**	2.5 معامل الانحدار الخطي البسيط
5.3 Multiple Linear Regression Coefficient	**117**	5. 3 معامل الانحدار الخطي المتعدد
5.4 Exercices	**122**	4.5 تمارين
Appendices	**126**	**ملاحق**
CONCLUSION	**130**	**الخاتمة**
Reference List	**135**	**قائمة المراجع**

المقدمة Preface

في عصر التدفق المستمر للبيانات والمعلومات، أصبحت القدرة على فهم الظواهر وتحليلها بدقة وعمق ضرورة أساسية لأي بحث علمي يهدف إلى الجدية والتميز. أهمية الإحصاء لم تعد مقتصرة على جمع البيانات ووصفها، بل تمتد إلى مجال حيوي يساعدنا في تجاوز حدود العينات لاستخلاص خصائص المجتمع الأصلي، واختبار الفرضيات، واتخاذ قرارات مبنية على أدلة رقمية موثوقة. هذا المجال، المعروف بالإحصاء الاستدلالي، يمثل الركيزة الأساسية للبحث الكمي وأحد الأدوات الأكثر أهمية للباحثين في العلوم الطبيعية والاجتماعية والإنسانية على حد سواء.

In an era of continuous data and information flow, the ability to understand and analyze phenomena accurately and thoroughly has become essential for any scientific research that aspires to rigor and excellence. The importance of statistics is no longer limited to data collection and description; it extends to a vital field that helps us transcend the limitations of samples to infer the characteristics of the original population, test hypotheses, and make decisions based on reliable numerical evidence. This field, known as inferential statistics, represents the cornerstone of quantitative research and one of the most important tools for researchers in the natural, social, and human sciences alike.

تقدم هذه المطبوعة مقدمة صارمة و بنيوية لنظرية الاحصاء الاستدلالي، مصممة خصيصا لطلبة العلوم الاجتماعية .نبدأ مسلمات الاحصاء الاستدلالي الاساسية ، التي ترسخ إطارا رياضيا متينا للتفكيرفي مسألة اتخاذ القرار واختيار المنهجية الاحصائية ، ثم ننتقل عبر مفاهيم رئيسية ذات الصلة بالعلوم الاجتماعية تشمل المتغيرات والتوزيعات وقوة الاختبار ومستويات الدلالة.

This publication provides a rigorous and structural introduction to inferential statistics theory, specifically designed for social science students. We begin with the basic postulates of inferential statistics, which establish a solid mathematical framework for thinking about decision-making and the selection of statistical methodology, and then move through key concepts relevant to the social sciences, including variables, distributions, test power, and levels of significance.

محور مركزي في هذه الدروس هو نتعلم كيفية اختيار الاختبار الإحصائي الملائم بناءً على طبيعة بياناتنا وأسئلة بحثنا، وكيفية تفسير النتائج واتخاذ القرارا بناءً عليها. كما نتعمق في أدوات التمثيل الأولي للبيانات عبر الجداول التكرارية بمختلف أشكالها، التي تمثل البوابة الأولى لفهم طبيعة العينة.

A central focus of these lessons is learning how to choose the appropriate statistical test based on the nature of our data and research questions, and how to interpret the results and make decisions accordingly. We also delve into the tools for initial data representation through various forms of frequency tables, which represent the first step towards understanding the nature of the sample.

وعلى امتداد الدروس، نسعى الى تقديم مجموعة من أهم أدوات التحليل، بدءاً من مقاييس الارتباط المختلفة (بيرسون، سبيرمان، كيندال) التي تكشف عن قوة واتجاه العلاقات بين المتغيرات، ووصولاً إلى اختبار كاي-2 الرائد في تحليل البيانات الوصفية وفحص الاستقلالية وحسن المطابقة، مروراً بنماذج الانحدار الخطي البسيط والمتعدد التي تسمح لنا بالتنبؤ والتقدير.

Throughout the lessons, we aim to present a range of the most important analytical tools, starting with different correlation measures (Pearson, Spearman, Kendall) that reveal the strength and direction of

relationships between variables, and culminating in the pioneering chi-two test for analyzing metadata and examining independence and goodness of fit , moving on to simple and multiple linear regression models that allow us to predict and estimate.

في الأخير، نسعى إلى تمكين الطالب من تحويل البيانات الخام إلى نتائج ذات معنى، واستخلاص استدلالات سليمة تدعم صرح المعرفة وتسهم في الإجابة على أسئلة البحث بثقة وموثوقية.

Finaly,we aim to empower students to transform raw data into meaningful results and draw sound inferences that contribute to the edifice of knowledge and help answer research questions with confidence and reliability.

الفصل الأول Chapter 1

الإطار المفاهيمي للاحصاء الاستدلالي

Concept framework of inferential statistics

تمهيد preface

يُمثل الإحصاء الاستدلالي الركيزة الأساسية للبحث العلمي الكمي، حيث يتيح للباحث تجاوز مجرد وصف البيانات إلى استنتاج خصائص المجتمعات الإحصائية الأكبر بناءً على عينات محدودة. فالسؤال المطروح إذا كان الإحصاء الوصفي يهتم بـ "ماذا حدث؟"، فإن الاستدلالي يجيب على سؤال "ماذا يعني هذا؟" و"هل يمكن تعميم النتائج؟"

Inferential statistics represents the cornerstone of quantitative scientific research, enabling researchers to move beyond simply describing data to inferring the characteristics of larger statistical populations based on limited samples. While descriptive statistics focuses on "What happened?", inferential statistics answers "What does this mean?" and "Can the results be generalized ?

لتسهيل استيعاب الفصول الموالية، سيركز هذا الفصل على الأسس النظرية والعملية للإحصاء الاستدلالي التي تشكل الأرضية للتطبيقات والاختبارات الإحصائية

To facilitate understanding of the following chapters, This chapter will focus on the theoretical and practical foundations of inferential statistics, which form the basis for statistical applications and tests.

1.1 اهمية الاحصاء الاستدلالي في البحث العلمي

1.1 The importanceof inferential statistics in scientific research

الإحصاء الاستدلالي في البحث العلمي يمثل أداة أساسية لتحويل الظواهر المعقدة إلى بيانات قابلة للقياس والتحليل. يسهم الإحصاء في جميع مراحل البحث بدءاً من التصميم وجمع البيانات وصولاً إلى التحليل والاستخلاص. بما يتميز به من أدوات رياضية دقيقة، يتيح للباحث اختبار الفرضيات بشكل موضوعي ويُساعد على التمييز بين الأنماط الحقيقية والصدف العشوائية. في عصر البيانات الضخمة، يصبح الإحصاء وسيلة حيوية لاستخلاص المعرفة من الكم الهائل من المعلومات

Inferential statistics in scientific research is a fundamental tool for transforming complex phenomena into measurable and analyzable data. Statistics contributes to all stages of research, from design and data collection to analysis and conclusion. With its precise mathematical tools, it allows researchers to test hypotheses objectively and helps distinguish between genuine patterns and random occurrences. In the age of big data, statistics has become a vital means of extracting knowledge from vast amounts of information.

2.1علاقة الاحصاء الاستدلالي بالعلوم الاجتماعية

1.2 The relationship of inferential statistics to the social sciences

في مجال العلوم الاجتماعية، الإحصاء الاستدلالي يمثل رابط أساسياً بين النظرية والواقع؛

حيث يحوّل المفاهيم المجردة إلى متغيرات قابلة للدراسة كما يُمكن الباحثين من استخدام مقاييس كمية لتحليل الظواهر البشرية والاجتماعية المعقدة. يمكن ملاحظة تطبيقات الإحصاء في عدة مجالات، مثل علم الاجتماع لدراسة الطبقات الاجتماعية والهجرة، وعلم النفس لقياس السمات الشخصية وتقييم أساليب العلاج، والاقتصاد لتحليل الأسواق وتوزيع الدخل، والعلوم السياسية لدراسة الانتخابات والرأي العام

In the social sciences, inferential statistics serves as a crucial link between theory and practice, transforming abstract concepts into measurable variables. It enables researchers to use quantitative measures to analyze complex human and social phenomena. Applications of statistics can be observed in various fields, such as sociology for studying social classes and migration, psychology for measuring personality traits and evaluating treatment methods, economics for analyzing markets and income distribution, and political science for studying elections and public opinion.

بالنهاية، الإحصاء الاستدلالي يتعدى كونه مجرد أداة رياضية؛ فهو منهجية تفكير تُمكن الباحثين في العلوم الاجتماعية من الانتقال من مجرد الوصف إلى التحليل العميق. يوفر الإحصاء إمكانية الاستدلال الموضوعي بعيداً عن الاعتماد الكامل على الحدس، كما يسمح بالتعميم من العينات الصغيرة إلى المجتمعات الكبيرة لاتخاذ قرارات مستنيرة تعتمد على أدلة كمية واضحة. هذه المساهمات تصبّ في تحسين السياسات والبرامج الاجتماعية بشكل ملموس ومدروس.

Finaly,inferential statistics is more than just a mathematical tool; it is a methodology that enables social scientists to move from mere description to in-depth analysis. Statistics provides the possibility of objective inference, moving beyond complete reliance on intuition, and allows for generalization from small samples to large populations, enabling informed decisions based on clear quantitative evidence. These contributions lead to tangible and well-considered improvements in social policies and programs.

3-1 المفاهيم الاساسية في الاحصاء

1-3Basic concepts in statistics

الوحدة الإحصائية: هي العنصر الأساسي الذي يشكل المجتمع الإحصائي.

Statistical Unit: This is the basic element that population constitutes the statistical

المجتمع الإحصائي :لا يُقصد بالمجتمع هنا معناه الشائع، بل هو في علم الإحصاء مجموعة من القياسات أو المشاهدات المرتبطة بمجموعة من الوحدات الإحصائية .بالتالي، المجتمع الإحصائي يشمل جميع العناصر أو الأفراد الذين يُركَّز عليهم في دراسة معينة .بمعنى آخر، هو جميع المكونات ذات الصلة بموضوع البحث، مثل طلاب جامعة معينة، سكان منطقة معينة،.... إلخ

Statistical Population: The term "population" is not the commonly used one. In the dictionary of statistics, a population is a set of observations

and measurements pertaining to a group of elements called statistical units. Therefore, it is the group of elements or individuals that are the focus of a particular study. In other words, it is all the elements related to the research topic. The study population might be the students of a specific university, the residents of a particular ,...etc

العينة الإحصائية: هي جزء أصغر مأخوذ من المجتمع الإحصائي محل الدراسة. يتم الاعتماد على العينة لإجراء الدراسة عندما يكون تطبيقها على كامل المجتمع أمراً صعباً.

Statistical Sample: This is a smaller portion taken from the statistical population under study. A sample is used to conduct a study when applying it to the entire population is impractical

الخاصية المدروسة (الميزة): هي الظاهرة التي يهتم الباحث بفهمها وتحليلها، وتعتبر الرابط المشترك بين جميع عناصر المجتمع أو العيّنة. مثال على ذلك: أطوال الطلاب. وتنقسم إلى نوعين:

The Investigated Characteristic(Feature) : This is the phenomenon that the researcher is interested in understanding and analyzing. It is considered the common link between all elements of the population or sample. An example is the height of students. Characteristics are divided into two types:

ا) الخاصية الكمية: عندما تُعبّر الخاصية عن قيم عددية وقابلة للعد، حيث يُطلق عليها اسم متغير كمي. ويمكن أن يكون هذا المتغير إما منفصلاً أو متصلاً.

A) Quantitative Characteristic :When a characteristic expresses numerical and countable values, it is called a quantitative variable. This variable can be either discrete or continuous.

ب) الخاصية النوعية: عندما لا تظهر الخاصية في شكل قيَمي، أي غير قابلة للعد، وتُطلق عليها حينها اسم متغير نوعي.

b) Qualitative Characteristic: When a characteristic does not appear in the form of values, i.e., it is uncountable, it is then called a qualitative variable.

المتغير الإحصائي: يعكس القيم المختلفة التي يمكن أن تأخذها الخاصية المدروسة. المتغير الإحصائي الكمي يمكن أن يكون منفصلاً أو متصلاً، بينما يكون في حالة المتغير النوعي عبارة عن توصيفات ورموز.

Statistical Variables: These reflect the different values that the studied characteristic can have. Quantitative statistical variables can be discrete or continuous, while qualitative variables consist of descriptions and symbols.

3.1 1-أنواع البيانات الإحصائية

1.3.1Types of Statistical Data

كلما كانت عملية جمع البيانات دقيقة، زادت ثقة الباحث بها وارتفعت جودة النتائج المستخلصة. ومع ذلك، فإن وجود أخطاء خلال جمع البيانات يؤدي إلى بطلان التحليل. بناءً على طبيعة الظاهرة المدروسة، تصنّف البيانات إلى:

The more accurate the data collection process, the greater the researcher's confidence in it and the higher the quality of the resulting findings. However, errors during data collection invalidate the analysis. Based on the nature of the phenomenon under study, data is classified into:

بيانات نوعية (كيفية أو وصفية)
Qualitative (or descriptive) data

تُنتج هذه البيانات عندما تكون الخاصية المدروسة غير قابلة للقياس العددي بل تعتمد على التصنيفات أو الفئات مثل الجنس أو الحالة الاجتماعية أو المهنية. ويُستخدم لقياسها عدة أساليب، منها:

This data is produced when the studied characteristic is not numerically measurable but rather relies on classifications or categories such as gender, marital status, or occupation. Several methods are used to measure it, including :

أ) **التدرج الاسمي:** يعتمد هذا المقياس على تصنيف العناصر بناءً على النوع فقط دون الكمية. تُمنح الأرقام لهوية العناصر ولكنها لا تدل على قيم كمية. مثال: استخدام الرقم **0** للجنس الذكر و**1** للجنس الأنثى، حيث أن هذه الأرقام لا تعكس دلالات عددية فعلية.

A) Nominal Gradation: This scale relies on classifying elements based solely on gender, not quantity. Numbers are assigned to identify the elements but do not represent quantitative values For example, using the number **1** for female and male **0** as these numbers do not reflect actual numerical significance.

ب) **التدرج الترتيبي:** يصنّف العناصر وفق ترتيب معين يُظهر تفضيلات أو مستويات مقارنةً بالتصنيف الاسمي، مثل تصنيف أداء الطلاب من ممتاز إلى ضعيف أو ترتيب التوافق مع سؤال ما مثل موافق جداً، موافق، متردد، لا أوافق.

b) Ordinal Ranking: This classifies elements in a specific order that reflects preferences or levels compared to nominal rankings. Examples include classifying student performance from excellent to poor, or ranking agreement with a question such as strongly agree, agree, undecided, and disagree.

بيانات كمية (عددية)

Quantitative (Numerical) Data

تُجمع هذه البيانات عندما تكون الخاصية قيد الدراسة قابلة للقياس أو العد. أمثلة على ذلك علامات الطلاب في الاختبارات، كميات الإنتاج، وعدد الموظفين. وتنقسم إلى:

This data is collected when the characteristic under study is measurable or countable. Examples include student test scores, production quantities, and the number of employees. It is divided into :

أ). <u>الكمي المنفصل:</u> يمثّل قيماً فردية غير قابلة للتجزئة مثل الأعداد الصحيحة.

<u>A)Discrete Quantitative</u>: This represents individual, as whole numbers indivisible values such.

ب). <u>الكمي المتصل:</u> يمثّل قيماً متعددة يمكن تقسيمها ووضعها في نطاقات أو مجالات.

<u>B)Continuous Quantitative:</u> This represents multiple values that can be divided and placed into ranges or domains.

1-4 الخطوات الاحصائية

1-4 Statistical steps

تعد الطريقة الإحصائية أداة مرنة تُستخدم في العديد من المجالات والتخصصات، مثل العلوم الاجتماعية والاقتصادية وغيرها. وقبل الخوض في تفاصيل هذه الطريقة، يمكن تسليط الضوء على أبرز أساليب جمع البيانات الإحصائية، ومنها:

Statistical methods are a flexible tool used in many fields and disciplines, such as social sciences, economics, and others. Before delving into the details of these methods, we can highlight the most prominent methods of collecting statistical data, including:

"المقابلة الشخصية" وتتم من خلال لقاء أفراد العينة ومناقشتهم حول موضوع الدراسة للحصول على معلومات دقيقة ومباشرة كذلك **"الاستبيان"** و الذي يتمثل في طرح مجموعة من الأسئلة على الأفراد بهدف التحقق من الفرضيات المتعلقة بالمشكلة قيد الدراسة. بالاظافة الى **"الملاحظة المباشرة "**والتي تُعتمد عندما تكون صعوبة في تحديد أفراد العينة، كما هو الحال في دراسة حركة السير عند مفترقات الطرق.

which involve meeting with ,"**Personal interviews**"
members of the sample and discussing the study topic with them to obtain accurate and direct information
which involve posing a set of ,"**Questionnaires**"
questions to individuals to test hypotheses related to the problem under study

"**Direct observation**"which is used when it is ,
difficult to identify members of the sample, as in studying traffic flow at intersections .

تهدف الطريقة الإحصائية إلى تحقيق عدة أهداف رئيسية منها:

Statistical methods aim to achieve several main objectives, including:

1.جمع البيانات والمعلومات : تسجيل الظواهر المختلفة بشكل رقمي وتصنيفها ضمن جداول منظمة وتمثيلها بالرسم البياني (الإحصاء الوصفي).

1-Data and Information Collection Recording various phenomena numerically, classifying them in organized tables, and representing them graphically (descriptive (statistics.

2. تحليل البيانات واستخلاص النتائج: استخدام البيانات لاتخاذ قرارات علمية ومدروسة (الإحصاء الاستدلالي)

2. Data Analysis and Drawing Conclusions: Using data to make informed and scientific decisions (inferential statistics).

3.مقارنة الظواهر والعلاقات بينها: لفهم طبيعتها والتعرف على القوانين التي تحكمها.

3.Comparing Phenomena and the Relationships Between :Them To understand their nature and identify the laws that govern them.

تمر الطريقة الإحصائية عبر مراحل محددة تشمل:

The statistical method involves specific stages, including:

- تعريف الظاهرة المطلوب دراستها

- Defining the phenomenon to be studied

- تحديد المجتمع المستهدف

-Identifying the target population

- جمع البيانات والمعلومات الإحصائية

-Collecting statistical data and information

- فرز وتنظيم البيانات

-Sorting and organizing the data.

- عرض البيانات في شكل جداول ورسوم بيانية

-Presenting the data in the form of tables and graphs

- تحليل وتفسير النتائج المستخلصة

-Analyzing and interpreting the results

تساعد هذه المراحل على تقديم فهم شامل ودقيق للظواهر المدروسة، مما يسهم في اتخاذ قرارات مبنية على أسس علمية واضحة

These stages help provide a comprehensive and accurate understanding of the phenomena under study, on contributing to informed decision-making based clear scientific principles

5.1 الجداول التكرارية والتمثيلات البيانية
1.5Frequency tables and graphs

بعد جمع البيانات يتوجب على الباحث ان ينظمها في جداول تكرارية وعرضها بيانيا وبالتالي نستعرض في هذا الجزء كيفية بناء جدول تكراري بسيط وجدول تكراري مركب وكذلك العرض البياني

After collecting the data, the researcher must organize it into frequency tables and present it

graphically. Therefore, in this section, we will review how to construct a simple frequency table, a complex frequency table, and the graphical presentation

إن إنشاء الجداول التكرارية البسيطة والمركبة يبدأ بتفريغ البيانات الخام من الاستمارات، حيث يعتمد الجدول البسيط على عمودين أساسيين: الأول يحتوي على القيم أو الفئات، والثاني يسجل التكرارات المطلقة، أي عدد مرات ظهور كل قيمة. يتم إضافة تعليق مختصر يوضح أكثر القيم شيوعاً أو يبين كيفية توزيعها. أما بالنسبة للجدول المركب أو التكراري المتجمع، فيتضمن عمودين إضافيين يعرضان التكرارات المجمعة الصاعدة والنازلة، مما يساعد في فهم تراكم التكرارات وتحليل إجمالي القيم الأقل أو الأكثر من حد معين. الهدف من ذلك هو تسهيل دراسة توزيع البيانات واكتشاف الأنماط بطريقة أكثر شمولاً.

Creating simple and compound frequency tables begins with extracting the raw data from the forms. A simple table relies on two main columns: the first contains the values or categories, and the second records the absolute frequencies, i.e., the number of times each value appears. A brief comment is added explaining the most common values or showing how they are distributed. A compound or cumulative frequency table includes two additional columns displaying the cumulative frequencies, both ascending and descending. This helps in understanding the accumulation of frequencies and analyzing the total of values that are below or above a certain threshold. The goal is to facilitate the study of data distribution and the discovery of patterns in a more comprehensive way

1.5.1 بناء وتحليل الجداول التكرارية البسيطة: من تفريغ البيانات إلى تفسير النتائج

1.5.1Construction and Analysis of Simple Frequency Tables: From Data Entry to Results Interpretation

لبناء جول تكراري بسيط (تفريغ استمارة) والتعليق عليه نتبع الخطوات التالية:

To construct a simple frequency table (form entry) and annotate it, follow these steps:

-تفريغ البيانات: بعد جمع الاستمارات، نقوم بإنشاء جدول بعمودين: "القيم "، و "التكرار المطلق" **(**عدد المرات)

- Data Entry: After collecting the forms, create a table with two columns: “Values” or categories and “Absolute Frequency”(number of times).

- ادخال البيانات : تسجيل كل قيمة أو فئة في عمود القيم ، ونقوم بعدّ كم مرة ظهرت كل قيمة في عمود التكرارات المقابل.

- Data extraction: Record each value or category in the Values column and count how many times each value appears in the corresponding Frequency column.

- **التعليق :** التعليق على أكثر القيم أو الفئات شيوعا ، (أعلى تكرار) و أقلها ، وحجم العينة ، أي ملاحظات هامة.

- Commentary: Comment on the most(frequent highest)and lowest frequency values or categories, the sample size, and any other important observations.

Example **مثال**

البيانات التالية توضح المستوى الدراسي لعشرين عامل:

The following data shows the educational level of twenty workers:

متوسط، ابتدائي، ابتدائي، ثانوي، جامعي، جامعي ، ابتدائي، متوسط، متوسط،، ابتدائي ، جامعي، ابتدائي، ابتدائي، ابتدائي، جامعي، ابتدائي، جامعي ، ابتدائي، جامعي، متوسط

Intermediate, Primary, Primary, Secondary, University, University, Primary, Intermediate, Intermediate, University, Primary, Primary, Primary, University, Primary, University, Intermediate, Primary.

-اعرض هذه البيانات في جدول تكراري

- Display this data in a frequency table

-اعرض هذه البيانات بيانيا

-Display this data graphically

العرض الجدولي للبيانات presentation table Data

التكرارات repetitions	educational level المستوى التعليمي
4	Intermediate متوسط
9	Primary ابتدائي
1	Secondary ثانوي
6	University جامعي
20	Sum المجموع

التعليق: يوضح الجدول ان المستوى الابتدائي هو المستوى الاكثر شيوعا للعمال يليه المستوى الجامعي ثم المستوى المتوسط و في الاخير المستوى الثانوي

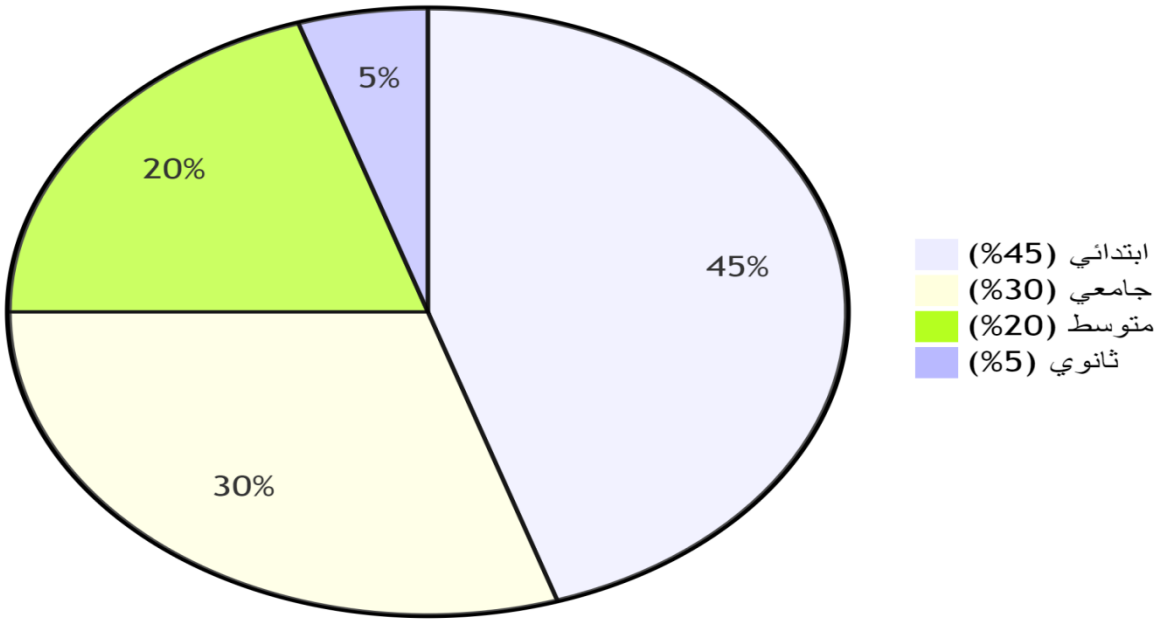
Comment: The table shows that the primary level is the most common level for workers, followed by the university level, then the intermediate level, and finally the secondary level.

العرض البياني للبيانات Data graphical presentation

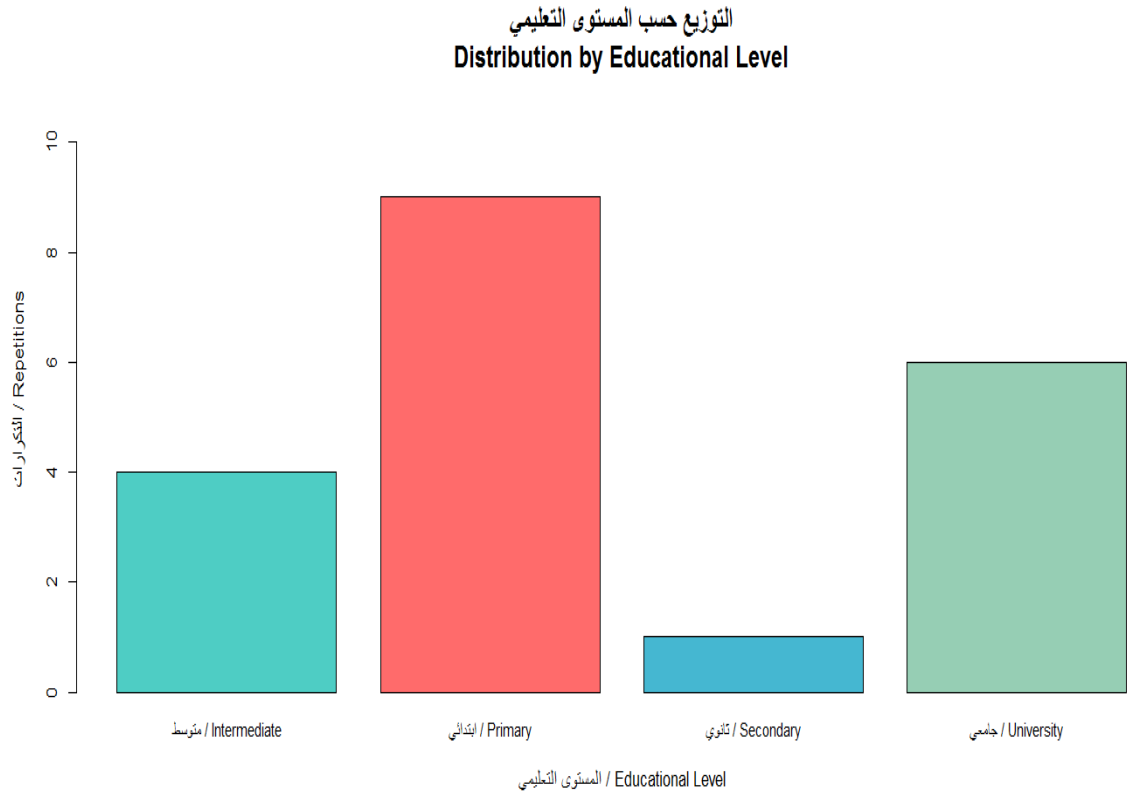
مخطط دائري : عرض النسب المئوية

Pie chart: displaying percentages

$$\text{الزاوية} = 360 * \frac{\text{التكرار}}{\text{مجموع التكرارات}}$$



Bar Chart مخطط أعمدة



1.5.2 تصميم وتفسير الجداول التكرارية المركبة: إجراءات البناء واستخلاص الدلالات

1-5-2 Design and Interpretation of Complex Frequency Tables: Construction Procedures and Inference Extraction

الجداول التكرارية المركبة هي جداول تعرض توزيع متغيرين أو أكثر، وتستخدم لتحليل العلاقات بين المتغيرات. و تنقسم الى عدة انواع منها :

Compound frequency tables are tables that display the distribution of two or more variables and are used to analyze the relationships between variables. They are divided into several types, including :

جدول ثنائي (2×2): متغيرين كل منهما له فئتين

table(2x2) Two variables, each with two categories

جدول متعدد الأبعاد: أكثر من متغيرين

Multidimensional table: More than two variables

لبناء جول تكراري مركب (تفريغ استمارة) والتعليق عليه نتبع الخطوات التالية:

To construct a composite frequency table)form entry(and annotate it, we follow these steps:

يبنى الجول البسيط اولا، ثم نضيف اليه اعمدة جديدة:

First, we build the simple table, then we add new columns to it:

1. **التكرار المجمع الصاعد:** نضيف عموداً "التكرار المجمع الصاعد (أقل من/أو يساوي)".

1. **Ascending cumulative frequency** we add a column Ascending cumulative frequency “less than/equal to”

الصف الاول: تكرار المطلق الأول

Roll 1 :First absolute frequency

الصف الثاني : (تكرار الصف الاول + تكرار الصف الثاني)
وهكذا، لتجميع التكرارات من الاسفل للأعلى

Roll 2 = (Frequency of Row 1 + Frequency of Row 2)
Continue summing frequencies from bottom to top.

2التكرار المجمع النازل : نضيف عموداً "التكرار المجمع النازل (أكبر من/أو يساوي)"

2.Descending cumulative frequency: we add a column " Descending cumulative frequency greater(than/equal to)"

الصف الأول = مجموع كل التكرارات

Row1=Sum of all frequencies

الصف الثاني = (مجموع التكرارات - تكرار الصف الأول)
وهكذا، لتجميع التكرارات من الأعلى للأسفل.

Row2=(Sum of frequencies -Frequency of Row)
Continue summing frequencies from top to bottom.

3.التعليق: نشرح ما تعنيه الأرقام المجمعة

3.Comment we explain whatthe summed numbers me an.

مثال : 1 جدول ثنائي 2x2 **Example 1** 2x2 table

البيانات التالية تمثل علامات مقياس علم الاجتماع لطلبة السنة الاولى العلوم الاجتماعية ل م د

The following data represents the scores of the sociology course for first-year social science students (LMD system).

11, 13, 14, 13, 17, 12, 13, 14, 15, 11,12, 14, 12, 13, 12, 14, 13, 12, 14, 13, 13, 14, 11, 13, 17, 13, 11, 17, 13, 15, 14, 12, 13, 15, 11, 12, 13, 14, 12, 13

المطلوب **Questions**

1-كون توزيع تكراري لعلامات الطلبة

1 -Construct a frequency distribution of student scores.

2 - ماهو عدد الطلاب الحاصلين على أقل من 13 درجة ؟

2 -How many students scored less than 13?

3- ماهو عدد الطلاب الحاصلين على 12 أو أكثر؟

3-How many students scored 12 or more ?

4- اعرض البيانات بيانيا

4- Display the data graphically

Solution **الحل**

يلاحظ من البيانات المعطاة أن درجات الطلاب في مقياس علم الاجتماع متغير كمي و يمكن تشكيل البيانات في جدول تكراري كالآتي

The given data shows that students' scores on the sociology scale are a quantitative variable, and the data can be structured into a frequency table as follows:

الدرجات X Grades	**11**	**12**	**13**	**14**	**15**	**17**	**المجموع Sum**
التكرار n_i repetition	**5**	**7**	**13**	**7**	**4**	**4**	**40**

لايجاد عدد الطلبة الحاصلين على علامات اقل من 13 و 12 او اكثر نقوم بإنشاء توزيع تكراري لمجمع الصاعد و النازل

To find the number of students with scores below 13 and 12 and or higher,we create a frequency

distribution of the ascending and descending cumulative scores.

الدرجات X Grades	التكرار n_i repetition	التكرار التصاعدي (≤) Ascending repetition	التكرار النازل (≥) Descending repetition
11	5	5	40
12	7	5+7= 12	40-5= 35
13	13	12+13= 25	35-7= 28
14	7	25+7= 32	28-13= 15
15	4	32+4= 36	15-7= 8
17	4	36+4= 40	8-4=4
المجموع Sum	40		

2-عدد الطلبة الحاصلين على 13 من عمود التوزيع التكراري الصاعد عند العلامة 12 نجد **12 طالبا**

2-The number of students who scored less than 13 in the ascending frequency distribution column at the score of 12 is **12 students.**

3-عدد الطلبة الحاصلين على علامات 12 او اكثر من عمود التوزيع التكراري النازل عند العلامة 13 نجد **35 طالبا**

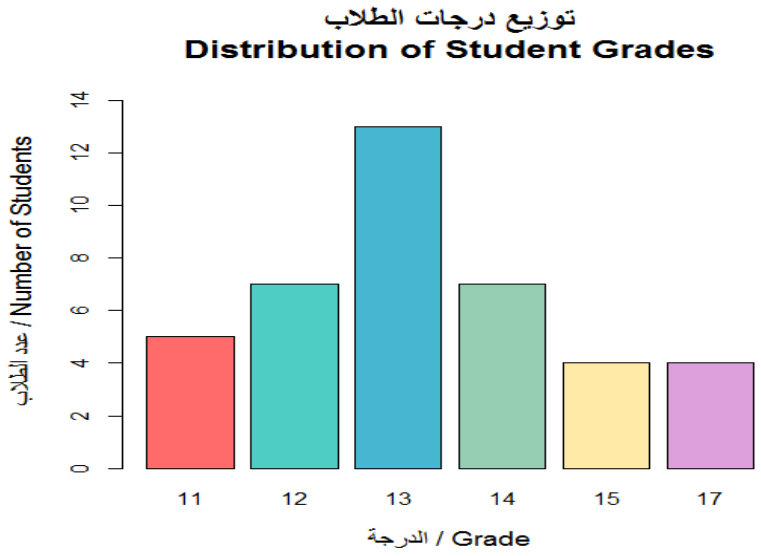
3 - The number of students who obtained marks 12 or higher from the descending frequency distribution co lumn at the mark 13 is **35 students**

4- عرض البيانات بيانيا

4- Present the data graphically

تمثيل البيانات بأعمدة بيانية كالتالي

Representing data using bar charts is as follows



مثال 2: (جدول متعدد الأبعاد)

Example 2 (multi-dimensional table)

البيانات التالية تمثل الجنس مع الرضا الوظيفي في شركة ما ل 20 عاملا :

The following data represents gender and job 20 satisfaction in a company for workers:

(ذكر، مرتفع)، (أنثى، متوسط)، (ذكر، منخفض)، (أنثى، مرتفع)، (أنثى، متوسط)، (ذكر، متوسط)، (أنثى، منخفض)، (ذكر، مرتفع)، (أنثى، مرتفع)، (ذكر، منخفض)، (أنثى، متوسط)، (أنثى، مرتفع)، (ذكر، متوسط)، (أنثى، منخفض)،

(ذكر، مرتفع)، (أنثى، متوسط)، (ذكر، متوسط)، (أنثى، مرتفع)، (ذكر، منخفض)، (أنثى، متوسط).

(Male, High) , (Female, Average) , (Male, Low) , (Female, High) , (Female, Average) , (Male, Average) , (Female, Low) , (Male, High) , (Female, High) ,

(Male, Low) , (Female, Average) , (Female, High) , (Male, Average) , (Female, Low), (Male, Average) , (Female, Average) , (Male, Average) , (Female, High),(Male, Low), (Female, Average).

-اعرض هذه البيانات في جدول تكراري

-Display this data in a frequency table

-اعرض هذه البيانات بيانيا

-Display this data graphically

اولا نقوم ببناء جدول تكراري الجنس والرضا الوظيفي بمتغيرين:

First, we construct a frequency table with two variables: gender and job satisfaction

الرضا \الجنس الوظيفي Gender / Job Satisfaction	**منخفض** **Low**	**متوسط** **medium**	**مرتفع** **high**	**المجموع** **Sum**
ذكر **Male**	**3**	**3**	**4**	**10**
انثى **Female**	**2**	**5**	**3**	**10**
Sum المجموع	**5**	**8**	**7**	**20**

ثانيا حساب النسب من ثلاث اتجاهات : الصف، العمود والإجمالي

Secondly, calculating percentages from three directions: row, column, and total.

نسبة الصف: نسبة كل خلية من مجموع صفها

Row percentage: The percentage of each cell out of its row

نسبة العمود : نسبة كل خلية من مجموع عمودها

Column percentage: The percentage of each cell out of its column

نسبةالإجمالي: نسبة كل خلية من المجموع الكلي

Total percentage: The percentage of each cell out of the total

لحساب نسبة الصف نقوم بالصيغة التالية

To calculate the row percentage, we use the following fo rmula

نسبة الصف للخلية = (قيمة الخلية ÷ مجموع الصف) × 100

Row percentage to cell=

100 ×)Cell value ÷ Row total)

For exemple **مثلا**

(Male, Low) cel ، : 3 ÷ 10 = 0.30 × 100 = %30

خلية (ذكر، منخفض)

أ) : 2 ÷ 10 = 0.20 × 100 = 20% ، (Female, Low) cell)
خلية (نثى، منخفض

To calculate the column percentage, we use the following formula

لحساب نسبة العمود نقوم بالصيغة التالية

نسبة العمود للخلية = (قيمة الخلية ÷ مجموع العمود) × 100

)Column percentage to cell = (Cell value ÷ Column total × 100

مثلا : For exemple

خلية(ذكر، منخفض) ، (Male, Low) cel : 3 ÷ 5 = 0.60 × 100 = 60%

خلية(أنثى، منخفض), (Female, Low) cell : 2 ÷ 5 = 0.40 × 100 = 40%

لحساب نسبة الإجمالي نقوم بالصيغة التالية

To calculate the percentage of the total, we use the following formula

نسبة الإجمالي = (قيمة الخلية ÷ المجموع الكلي) × 100

100 ×)Total percentage = (Cell value ÷ Total

for example مثلا

خلية(ذكر، منخفض) ، (Male, Low) cel: 3 ÷ 20 = 0.15 × 100 = 15%

وهكذا بفس الطريقة نحسب كل الخليات و نتحصل على النتائج في الجدول التالي :

We calculate this for all cells in the same way and obtain the results in the following table:

الرضا الوظيفي \الجنس Gender / Job Satisfaction	منخفض Low	متوسط medium	مرتفع high	النسبة /المجموع Percentage/ Sum
Male ذكر	**3**	**3**	**4**	**10/ 100%**
من صف الذكور % From the male rows	**30%**	**30%**	**40%**	**100%**
من عمود الرضا % From the satisfaction column	**60%**	**37.5%**	**57.1%**	حصة الذكور في كل مستوى رضا **Male share at each satisfaction level**
Female انثى	**2**	**5**	**3**	**10 / 100%**
من صف الإنا ث% From the female row	**20%**	**50%**	**30%**	**100%**
من عمود الرضا% From the satisfaction column	**40%**	**62.5%**	**42.9%**	حصة الإناث في كل مستوى رضا **Female share at each satisfaction level**
المجموع Sum	**5**	**8**	**7**	**20**
من المجموع % From the total	**25%**	**40%**	**35%**	**100%**

التعليق:

نلاحظ أن الذكور يتوزعون بشكل شبه متساوٍ عبر مستويات الرضا الثلاثة (30% منخفض، 30% متوسط، 40% مرتفع)، بينما تظهر الإناث تركيزاً واضحاً في فئة الرضا

المتوسط التي تشكل 50% منهن، مع انخفاض ملحوظ في نسبة الرضا المنخفض (20% فقط). هذا التباين في أنماط التوزيع يُلمح إلى اختلاف في العوامل المؤثرة على الرضا الوظيفي لدى كل جنس، حيث تبدو الإناث أكثر ميلاً نحو المستويات المتوسطة والمرتفعة، بينما يظهر الذكور انتشاراً أوسع عبر جميع المستويات.

Comment :

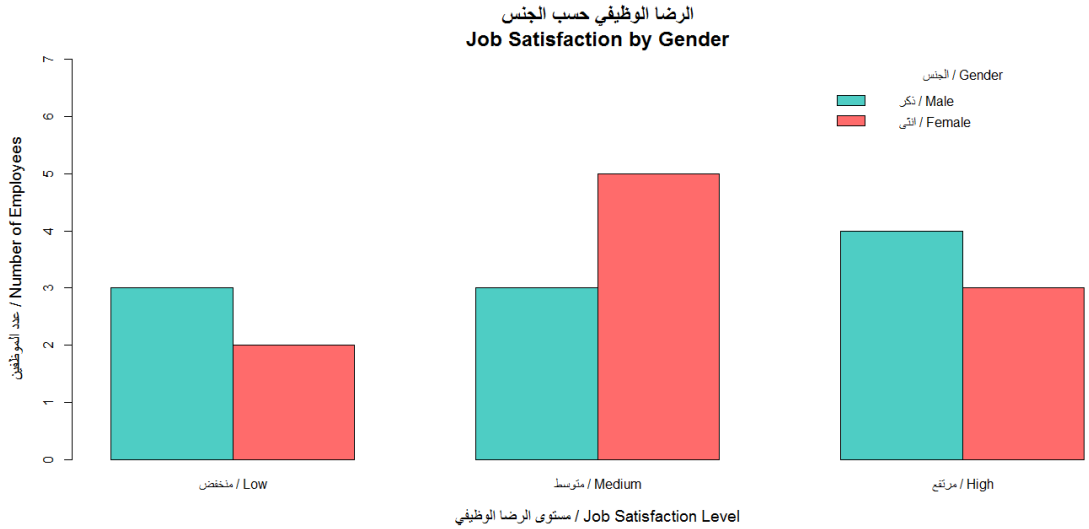
We note that males are distributed almost equally across the three satisfaction levels (30% low, 30% medium, 40% high).

females show a clear concentration in the medium of them, with %50 satisfaction category, comprising a notable decrease in the low satisfaction category .)%20 only) This variation in distribution patterns suggests a difference in the factors influencing job satisfaction for each gender, with females appearing more inclined towards the medium and high levels, while males exhibit a wider distribution across all levels.

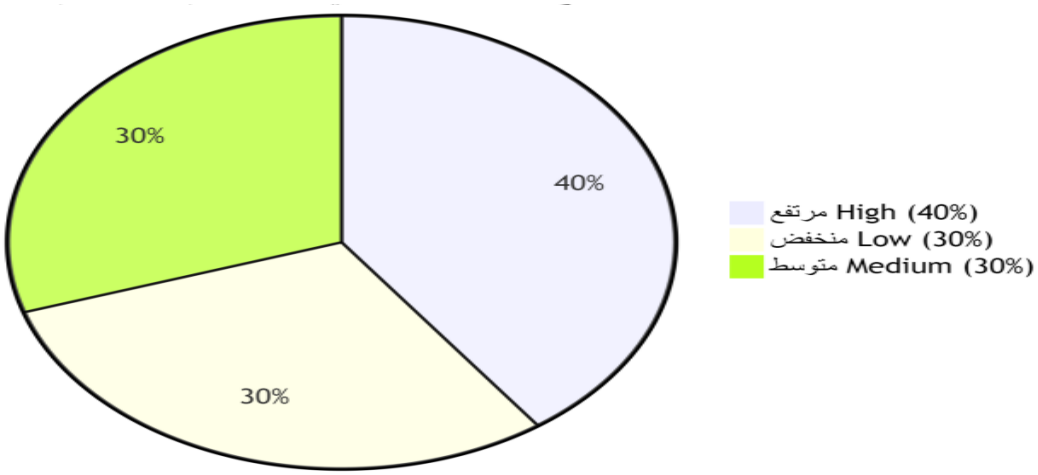
التمثيل البياني للبيانات

Data graphical representation

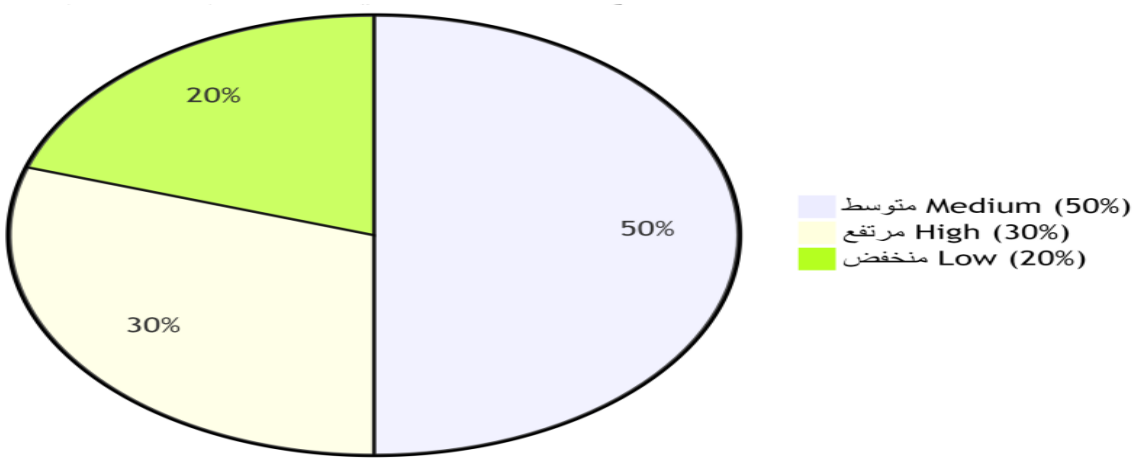
A. Grouped column diagram **أ.مخطط الاعمدة المجمعة**



ب. الدائرة البيانية b. The pie chart



توزيع الرضا الوظيفي للذكور Distribution of male job satisfaction



توزيع الرضا الوظيفي للإناث Distribution of female job satisfaction

تمارين Exercices

التمرين الاول **Exercice1**

اجب بصحيح ام خطأ مع تصحيح الخطأ إن وجد للعبارات التالية:

Answer True or False, correcting any errors:

المجتمع هو مجموعة جزئية من العينة.

- The population is a subset of the sample.

الفرد هو أصغر وحدة تحمل قيماً للمتغيرات المدروسة.

The individual is the smallest unit that holds values for the variables under study.

المتغير "التخصص" هو متغير كمي منفصل.

- The variable "specialization" is a discrete quantitative variable.

المتغير "الطول " هو كمي متصل.

The variable "height" is a continuous quantitative variable.

التمرين الثاني **Exercice 2**

على يدرس باحث موضوع "تأثير استعمال الذكاء الاصطناعي التحصيل الدراسي لطلبة جامعة بسكرة

A researcher is studying the topic of "The Impact of Artificial Intelligence on the Academic Achievement of Students at the University of Biskra

حدد المجتمع الإحصائي في هذه الدراسة؟و من هو الفرد؟

Define the statistical population in this study. What constitutes an individual?

اذكر متغيرين اثنين على الأقل يمكن للباحث قياسه، وحدد نوع كل منهما.

List at least two variables that the researcher can measure, and specify the type of each.

إذا أراد الباحث استقراء النتائج من عينة إلى المجتمع، فهل يحتاج إلى إحصاء وصفي أم استدلالي؟ ولماذا؟

If the researcher wants to extrapolate results from a sample to the population, does he need descriptive or inferential statistics? Why?

Exercice 3 **التمرين الثالث**

البيانات التالية تم جمعها من 15 طالباً في كلية العلوم الإجتماعية، حيث سُئل كل طالب عن:

The following data was collected from 20 students in the College of Social Sciences, where each student was asked about:

المستوى الدراسي (سنة أولى، سنة ثانية، سنة ثالثة) والرغبة في الالتحاق ببرنامج تدريب صيفي (نعم، لا، لم أحدد بعد) وكانت النتائج كالتالي

Academic level (first year, second year, third year) and desire to join a summer training program (yes, no, next, I have not yet determined) and the results were:

(أولى-نعم) - (أولى-نعم) - (أولى-لا) - (أولى-لم يحدد) - (أولى-نعم) - (ثالثة-نعم) - (ثانية-لا) - (ثانية-نعم) - (أولى-لم يحدد) - (أولى-نعم) - (ثانية-نعم) - (اولى-لا) - (ثانية-نعم) - (ثانية-نعم) (ثالثة-لم يحدد)

(First-Yes)- (First-Yes)- (First-No)- (First-not specified)- (First-Yes)- (Third-yes)- (Second-no)- (Second-yes)- (First-not specified)- (First-yes)-

(Second-yes)- (First-no)- (Second-yes)- (Second-yes)- (Third-not specified)

1 قم بتفريغ البيانات (أي عد التكرارات) في جدول تكراري مركب و التعليق عليه

1 Translate the data (i.e. count the occurrences) into a compound frequency table and annotate it

2مثل الاحصائية بالتمثيل البياني المناسب

2 Represent the statistics with an appropriate graphical representation.

التمرين الرابع **Exercice 4**

في دراسة حول البطالة لدى خريجي الجامعة بولاية بسكرة تم الحصول على الشهادات المتحصل عليها بعد التخرج كلتالي :

In a study on unemployment among university graduates in the state of Biskra, the certificates obtained after graduation were as follows:

ليسانس- ماستر- دكتوراه - ماستر- ليسانس- ماستر - ليسانس- دكتوراه- دكتوراه- ليسانس- ليسانس- ماستر - ماستر- ليسانس - ليسانس- ماستر- - ليسانس-ليسانس- دكتوراه- دكتوراه- ماستر- ليسانس ليسانس- دكتوراه - ماستر- ليسانس- ماستر- ليسانس- ماستر- ليسانس- ماستر

Bachelor - Master - Doctorate - Master - Bachelor - Master - Doctorate - Bachelor - Bachelor - Master - Bachelor - Doctorate - Bachelor - Bachelor - Master - Bachelor - Bachelor - Master - Doctorate - Doctorate - Master - Bachelor Bachelor - Doctor - Master - Bachelor - Master - Bachelor - Master - Bachelor - Bachelor – Master

1 ما نوع البيانات المتحصل عليها في الدراسة

1 What type of data is obtained in this study?

2 أنشئ جدول احصائي يوضح توزيع افراد العينة واحسب التكرار النسبي المئوية حسب الشهادة المتحصل عليها ؟

2 Create a statistical table showing the distribution of the sample members and calculate the

percentage frequency according to the certificate obtained?

3 مثل البيانات الإحصائية بالشكل المناسب لها ؟

3 Such as statistical data in the appropriate form for it?

Chapter 2 الفصل الثاني

الاسس النظرية والمعايير المنهجية لاختبار الفرضيات

Theortical Foundations and Methodolological Criteria for Hypothesies Testing

تمهيد Preface

يركز هذا الفصل على تزويد الباحث بالاسس المنهجية الضرورية لاتخاذ القرار الصحيح في اختيار الاختبار الاحصائي الملائم، فاختيار اختبار غيرمناسب ينتج عنه نتائج مغلوطة واستنتاجات بحثية غير دقيقة حيث يستند الاختيار الدقيق إلى ثلاثة عناصر أساسية:

طبيعة الفرضيات، خصائص العينة ومستوى قياس البيانات. وكبداية لهذا الفصل نتطرق الى اهم **التوزيعات الاحتمالية في الاحصاء الاستدلالي**

This chapter focuses on providing researchers with the necessary methodological foundations for making the right decision in selecting the appropriate statistical test. Choosing an unsuitable test leads to erroneous results and inaccurate research conclusions. Accurate selection is based on three fundamental elements: the nature of the hypotheses, the characteristics of the sample, and the level of data measurement. As a starting point for this chapter, we will discuss the most important **probability distributions in inferential statistics.**

1.2 التوزيعات الاحتمالية في الاحصاء الاستدلالي

2.1 Probability distributions in statistical inference

يعتمد اختيار الاختبار الإحصائي وتفسير نتائجه بشكل أساسي على افتراضات حول التوزيع الأساسي للبيانات وبالتالي تشكل التوزيعات الاحتمالية القاعدة الرياضية التي تُبنى عليها جميع الاستدلالات الإحصائية. فهي النماذج النظرية التي تصف كيفيةتوزيع القيم المحتملة لمتغير ما، وتحدد احتمالات حدوث النتائج المختلفة، وتتمثل هذه التوزيعات في:

The selection of a statistical test and the interpretation of its results depend primarily on assumptions about the underlying distribution of the data. Probability distributions, therefore, form the mathematical foundation

upon which all statistical inferences are built. They are theoretical models that describe how the possible values of a variable are distributed and determine the probabilities of different outcomes.These distributions include :

I. التوزيع الطبيعي I.The normal distribution

المعروف بالتوزيع المعتدل، يعد واحد من ابرز التوزيعات الاحتمالية المستمرة ، حيث يشكل الاساس الذي تستند إليه التطبيقات ،في الاستدلال الاحصائي ، بالاضافة الى العديد من ذلك ، فإن معظم التوزيعات الاحتمالية الاخرى سواءا كانت متقطعة او مستمرة ، يمكن ان تميل الى التوزيع الطبيعي عند توفر شروط معينة.

Also known as the moderate distribution, is one of the most prominent continuous probability distributions, as it forms the basis upon which many applications in statistical inference are based. In addition, most other probability distributions, whether discrete or continuous, can tend towards the normal distribution when certain conditions are met.

ونعبر عن ذلك رياضيا بالكتابة التالية $x_i \sim N(\mu_x, \sigma_x)$

This can be expressed mathematically as follows: $x_i \sim N(\mu_x, \sigma_x)$

حيث دالة الكثافة الاحتمالية لهذا التوزيع هي

Where the probability density function of this distribution is

$$f(x, \mu_x, \sigma_x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2{\sigma_x}^2}}$$

Where حيث

N: رمز التوزيع الطبيعي normal distribution symbol

X : متغير عشوائي random variable

μ_x متوسط المجتمع Population mean

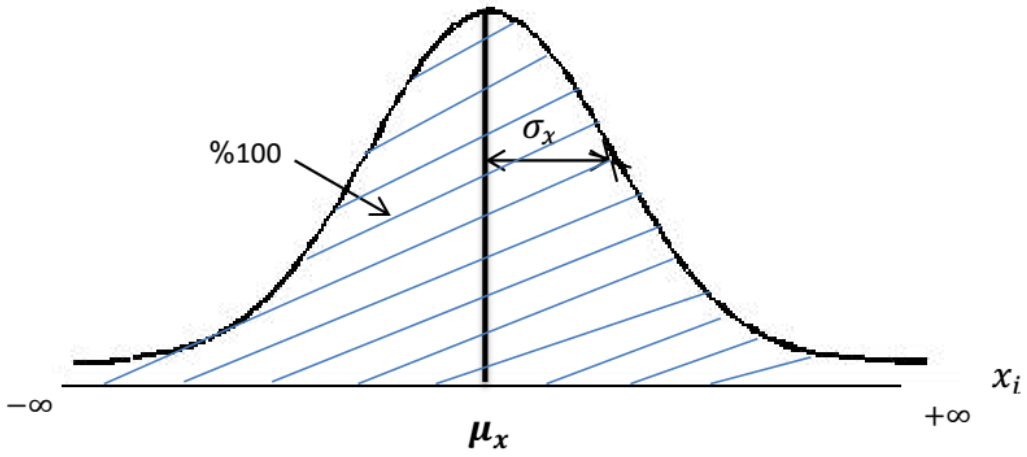
σ_x : الانحراف المعياري Standard deviation

$\pi = 3.14$ ثابت A constant

e=2.781 اساس اللوغاريتم النيبيري The basis of the Naberian logarithm

كما يأخذ التمثيل البياني كدالة كثافة احتمال التوزيع الطبيعي شكل جرسي متناظر بالنسبة للوسط الذي يوضحه الشكل التالي :

the graph representing the probability density of the nor-mal distribution also takes a bell-shaped form symmetricalto the mean, as illustrated in the following figure



يمكن استخلاص مجموعة من الخصائص التي تميز التوزيع الطبيعي من خلال الشكل :

A number of properties that characterize the normal distribution can be deduced from the figure:

1 . منحى التوزيع الطبيعي دائمًا يكون فوق المحور الأفقي ، مما يعني أنه موجب القيمة على الدوام.

1. The curve of the normal distribution is always above the horizontal axis, meaning it is always positive.

2 . نطاق المتغير العشوائي الذي يتبع التوزيع الطبيعي يمتد من$-\infty$ الى$+\infty$

2. The range of a random variable following the normal distribution extends $-\infty$ to $+\infty$.

3 . أي جزء من المساحة أسفل منحنى التوزيع الطبيعي يمثل احتمالًا معينًا، وبما أن مجموع الاحتمالات لأي توزيع احتمالي يساوي واحد، فإن المساحة الكلية تحت المنحنى تساوي واحد (أي 100%). وهذا يعني أن تكامل دالة الكثافة على كل مجال التعريف يساوي واحد اي$\int_{-\infty}^{+\infty} f(x) = 1$

3. Any part of the area under the curve of the normal distribution represents a particular probability. Since the sum of the probabilities of any probability distribution equals one, the total area under the curve equals one (i.e., 100%). This means that the integral of the density function over each domain of definition equals one i.e., $\int_{-\infty}^{+\infty} f(x) = 1$

4 . منحنى التوزيع الطبيعي ذو تناظر كامل حول المتوسط، مما يجعله متماثلًا بحيث يكون النصف الأيسر متساويًا تمامًا مع النصف الأيمن بالنسبة لخط المتوسط.

4. The curve of the normal distribution is perfectly symmetrical about the mean, making it symmetrical so that the left half is exactly equal to the right half with respect to the mean line.

Remark: **ملاحظة :**

ان حساب احتمال وقوع متغير عشوائي طبيعي ضمن مجال محدد المجال أي [a,b] يتم من خلال إيجاد نتيجة تكامل دالة كثافة التوزيع الطبيعي داخل هذا المجال بمعنى:

Calculating the probability of a normal random variable occurring within a given range is done [a,b] by finding the result of integrating the normal distribution density function within that range i.e,

$$\mathbf{p(a \le x \le b) = \int_a^b \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x{}^2}}}$$

نظرا لصعوبة حساب هذا الاحتمال بشكل مباشر، يتم اللجوء إلى توزيع آخر يُعرف بالتوزيع الطبيعي المعياري ، والذي يُعتبر حالة خاصة من التوزيع الطبيعي حيث يكون المتوسط الصفري و الانحراف المعياري يُساوي 1. ايμ_x=0 و σ_x =1

Because calculating this probability directly is difficult, another distribution known as the standard normal distribution is used. This is a special case of the normal distribution where the mean is zero and the standard deviation is i.e, σ_x =1 and μ_x=0

وبناء على ذلك، يتم حساب الاحتمال كالتالي :

Therefore, the probability is calculated as follows :

استبدال المتغير x_i **الذي يتبع التوزيع الطبيعي** $\mathbf{x_i \sim N(\mu_x, \sigma_x)}$

بمتغير اخر $\mathbf{Z_i}$ **يتبع التوزيع المعياري الطبيعي** $\mathbf{Z_i \sim N(0,1)}$ وفق العلاقة التالية

Replace the variable x_i that **follows the normal distribution** $x_i \sim N(\mu_x, \sigma_x)$ with another variable Z_i that **follows the standard normal distribution** $Z_i \sim N(0,1)$ according to the following relationship:

$$Z_i = \frac{x - \mu_x}{\sigma_x} \sim N(0,1)$$

مثال Example

تقطع ماكنة نجارة الحديد أخشابا من الحديد الصناعي بقياسات تتبع التوزيع الطبيعي بمعدل سمك 25 ملم و تباين قيمته 9 ملم أخذت عشوائيا لوحة من بين اللوحات المقطوعة للإختبار

A metalworking machine cuts pieces of industrialiron to dimensions that follow a normal distribution, with an average thickness of 25 mm and a variance of 10 mm. A piece is randomly selected from among the cut pieces for testing.

-اكتب التوزيع النظري لهذا المتغير (السمك) واحسب احتمال ان يكون سمكها اكبر من 35 ملم

-Write the theoretical distribution of this variable (thickness) and calculate the probability that its thickness will be greater than 35 mm.

- احسب احتمال أن يقع سمكها بين القيمتين 20 و40 ملم

-Calculate the probability that its thickness will fall bet ween the values of 20 and 40 mm.

الحل **Solution**

1-Write theorticale distribution

1-كتابة التوزيع النظري

لدينا we have

$$f(x, \mu_x, \sigma_x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

$$\mu_x = 25, \sigma_x^2 = 9$$

So اذن

$$f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-25)^2}{2(9)}} \quad \sim N(25,9)$$

2-حساب احتمال ان يكون السمك اكبر من 35 ملم

2-Calculate the probability that the thickness will be greater than 30 mm

$$P(X > 35) = 1 - P(X < 35)$$
$$= 1 - P\left(\frac{x-\mu_x}{\sigma_x} < \frac{35-25}{3}\right)$$
$$= 1 - P(Z < 3.33)$$
$$= 1 - F(3.33)$$

من جدول التوزيع الطبيعي المعياري $F(3.33) = 0.99$

(انظر الملحق 1)

from the standard normal distribution table $F(3.33) = 0.99$ (**see Appendix 1**)

$$P(X > 35) = 1 - 0.99 = 0.000434$$

3-حساب احتمال أن يقع سمكها بين القيمتين 20 و40 ملم

3- Calculate the probability that its thickness will fall betwe en the values of 20 and 40 mm.

$$P(20 < X < 40) = P\left(\frac{20-25}{3} < \frac{x-\mu_x}{\sigma_x} < \frac{40-25}{3}\right)$$
$$= P(-1.6667 < Z < 5)$$
$$= F(5) - F(-1.6667)$$
$$= F(5) - (1 - F(1.6667))$$

من جدول التوزيع الطبيعي المعياري $F(5)=1$ و $F(1.6667) = 0.9525$

from the standard normal distribution $F(1.6667) = 0.9525$ *and* $F(5)=1$

$$P(20 < X < 40) = 1-(1-0.9525)$$

$$P(20 < X < 40) = 0.9525$$

II- Student distribution II.توزيع ستودنت :

إذا كان لدينا متغير عشوائي **t** يتبع توزيع ستودنت فإن دالة كثافته تُعطى بالعلاقة التالية

If we have a random variable **t** that follows a Student's distribution, its density function is given by the following relation

$$\mathbf{f(t) = c\left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}}$$

Where حيث:

n تمثل حجم العينة و **v=n-1** حيث **v** تمثل درجة الحرية

n represents the sample size and **v = n-1** where **v** represents the degrees of freedom

c: عبارة عن ثابت It is a constant

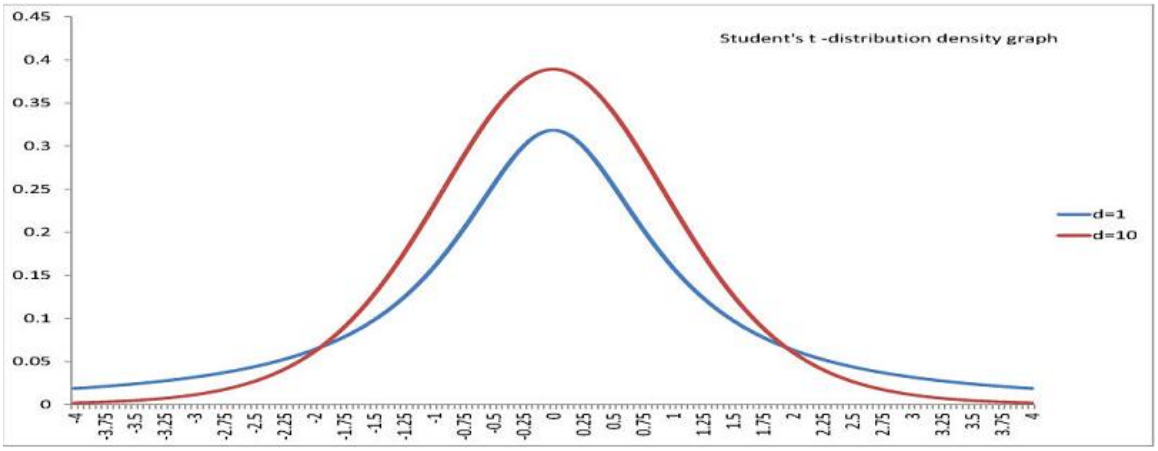
And, we write ونكتب

$$\mathbf{t \sim T_{\alpha.v}}$$

يمتلك التوزيع البياني لكثافة توزيع ستودنت شكلًا مشابهًا للتوزيع الطبيعي، ولكنه يكون أقل ارتفاعًا منه. يعتمد منحنى توزيع ستودنت بشكل رئيسي على درجات الحرية. حيث كلما زادت درجات الحرية اقترب شكل توزيع ستودنت أكثر من شكل التوزيع الطبيعي. يمكن توضيح هذه العلاقة في الشكل التالي:

The density distribution of the Student's distribution has a shape similar to the normal distribution, but at a lower elevation. The Student's distribution curve depends primarily on degrees of freedom; the higher the degrees of

freedom, the closer the Student's distribution resembles the normal distribution. This relationship can be illustrated in the following figure:



من خلال الملاحظة يتضح أن توزيع ستودنت له تشابه كبير مع التوزيع الطبيعي، ويشترك معه في العديد من الخصائص، منها:

Observation reveals that the Student's distribution is very similar to the normal distribution, sharing many characteristics, including:

1. مدى المتغير العشوائي t الذي يتبع توزيع ستودنت يمتد من

$-\infty$ إلى $+\infty$

1.The range of a random variable following the Student's distribution extends from $-\infty$ to $+\infty$

2 . التمثيل البياني لتوزيع ستودنت يأخذ شكلاً جرسياً، لكنه أقل ارتفاعاً من منحنى التوزيع الطبيعي، ويزداد ارتفاعه تدريجياً مع زيادة درجات الحرية.

2.The graph of the Student's distribution is bell-shaped, but it is lower than the curve of the normal distribution, and its height increases gradually with increasing degrees of freedom

3 . المساحة تحت المنحنى تساوي واحداً، مما يعني أنه عند تحديد قيمة معينة يمكن حساب احتمال المنطقة المحددة.

3.The area under the curve is equal to one, meaning that for a given value, the probability of that specific region can be calculated

4. توزيع ستودنت يعتبر متناظرا حول الوسط الحسابي وهو الصفر

4. The Student's distribution is symmetrical about the mean, which is zero

5 . بناءً على الخاصيتين الثالثة والرابعة، فإن القيمتين المتناظرتين (الإيجابية والسالبة) لهما احتمالات متساوية، أي أن مجموع هذه الاحتمالات يساوي واحدا

5.Based on the third and fourth characteristics, the two positive and negative) have equal) corresponding values probabilities; that is, the sum of these probabilities equals one

بالتالي ، نعد جدول توزيع ستيودنت خصيصا من أجل قراءة قيم معينة في مجال تعريفه وذلك بدلالة كل من: درجة الحرية

Nو احتمال معلوم **p** حيث $\mathbf{P(x \leq x_i)}$. تسمى هذه القيم بالقيم الجدولية $t_{p,n}$ مثلا

Therefore, we construct a Student's distribution table specifically to read certain values within its domain in terms of both the degrees of freedom **N** and a known probability **p**, where $\mathbf{P(x \leq x_i)}$. These values are called tabular values $t_{p,n}$

For example

* $t_{p,n} = t_{0.7,15} = 0.534 \Rightarrow \mathbf{P(T_{10} \leq 0.534) = 0.7}$

* $t_{p,n} = t_{0.5,20} = 0 \Rightarrow \mathbf{P(T_{20} \leq 0) = 0.5}$

* $t_{p,n} = t_{0.95,18} = 1.734 \Rightarrow \mathbf{P(T_{18} \leq 0.534) = 0.95}$

III. Chi 2 Disrtibution

III. توزيع كاى مربع

يعتبر من التوزيعات الاحتمالية المهمة ويرمز لتوزيع كاي مربع بالرمز $X^2_{a,v}$

It is considered one of the important and well-known probability distributions, and the chi-square distribution is symbolized by $X^2_{a,v}$

يتمتع بأهمية كبيرة في الدراسات الإحصائية،و يُستخدم في تحليل العينات وتقدير التباين وغير ذلك من التطبيقات الإحصائية. كما تعطى دالة كثافة التوزيع بالعلاقة التالية :

It is of great importance in statistical studies and is used in sample analysis, variance estimation, and other statistical applications. The distribution density function is given by the following relationship:

$$\mathbf{f(X^2) = c(X^2)^{\frac{v-2}{2}}\, e^{\frac{-X^2}{2}}}$$

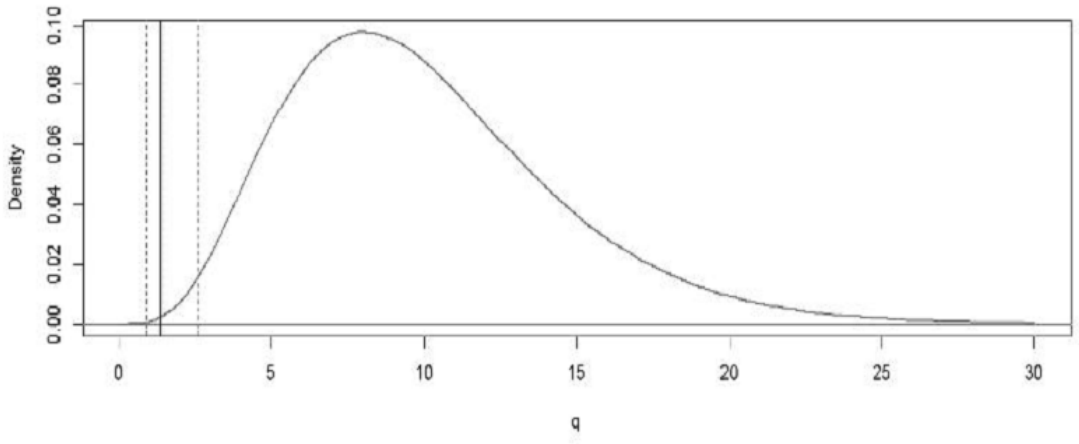
Where حيث :

v : درجة الحرية **v=(n-1) // n** حجم العينة ، **c** ثابت

v:degree of freedoom **v=(n-1)// n** sample size , **c** is canstant

يكون التمثيل البياني لدالة كثافة التوزيع على النحو التالي

The graph of the distribution density function is as follows:



ومن خلال التمثيل نلاحظ خصائص التوزيع وهي كالاتي

Through representation, we observe the characteristics of the distribution, which are as follows:

1-مدى المتغير الذي يتبع توزيع كاي مربع يبدأ من الصفر ويمتد إلى اللانهاية. وبناءً على ذلك، يكون التمثيل البياني لهذا التوزيع محصورًا في الربع الأول من المستوى.

1-The range of a variable following a chi-squared distribution starts at zero and extends to infinity. Therefore, the graph of this distribution is confined to the first quartile of the plane

2-يتميز التمثيل البياني لتوزيع كاي مربع بعدم التناظر، حيث يميل نحو اليمين. يكون مدى التوزيع محدودًا عند الصفر من الجهة اليسرى، وغير محدود من الجهة اليمن

2-The graph of a chi-squared distribution is characterized by asymmetry, as it slopes to the right. The range of the distribution is bounded at zero on the left side and unlimited on the right side

بالتالي، نعد جدول توزيع ستيودنت خصيصا من أجل قراءة قيم معينة في مجال تعريفه وذلك بدلالة كل من: درجة الحرية

Nو احتمال معلوم **p** حيث $\mathbf{P(x \leq } x_i\mathbf{)}$. تسمى هذه القيم بالقيم الجدولية $k_{p,n}^2$ مثلا

Therefore, we construct a Student's distribution table specifically to read certain values within its domain in terms of both the degrees of freedom **N** and a known probability **p**, where $\mathbf{P(x \leq } x_i\mathbf{)}$. These values are called tabular values $k^2_{p,n}$

For example

إذا كان **x** متغيرا عشوائيا يتبع توزيع كاي تربيع بدرجة حرية **n** تساوي **30**. أوجد القيم الجدولية التي تحقق:

If **x** is a random variable that follows a chi-squared distribution with **n** degrees of freedom equal to **30**, find the tabulated values that satisfy:

$$\mathbf{P}\left(x \leq k^2_{p,n}\right) = \mathbf{0.90}, \quad \mathbf{P}\left(x \leq k^2_{p,n}\right) = \mathbf{0.05}$$

اعتمادا على جدول توزيع كاي مربع (**الملحق 3**) نجد:

Based on the chi-squared distribution table(**Appendix 3**), we find:

$$\mathbf{P}\left(x \leq k^2_{p,n}\right) = \mathbf{0.90} \Rightarrow\Rightarrow \mathbf{P}\left(x \leq k^2_{0.90,30}\right) \rightarrow k^2_{0.90,30} = 40.256$$

$$\mathbf{P}\left(x \leq k^2_{p,n}\right) = \mathbf{0.05} \Rightarrow\Rightarrow \mathbf{P}\left(x \leq k^2_{0.05,30}\right) \rightarrow k^2_{0.05,30} = 18.493$$

V. Fisher Distribution — V. توزيع فيشر

يعتبر من التوزيعات الاحتمالية المهمة التي تُستخدم في الاستدلال الإحصائي، خاصة في الجزء المتعلق باختبار الفرضيات،

ويرمز له بالرمز $F_{v_1 v_2,n}$ ،و تعطى دالة كثافة توزيع فيشربالعلاقة التالية:

It is considered one of the important probability distributions used in statistical inference, especially in the hypothesis testing section, and is denoted by $F_{v_1 v_2,n}$. The density function of the Fisher distribution is given by the following formula

$$\mathbf{f(F)} = \frac{\mathbf{cF}^{\frac{\mathbf{v_1}-2}{2}}}{(\boldsymbol{v_2} + \boldsymbol{v_1 F})^{\frac{\boldsymbol{v_1}+\boldsymbol{v_2}}{2}}}$$

Where بحيث:

$\boldsymbol{v_1}$ درجة حرية البسط $\boldsymbol{v_1} = \boldsymbol{n_1}-1$ // $\boldsymbol{n_1}$ حجم العينة الاولى

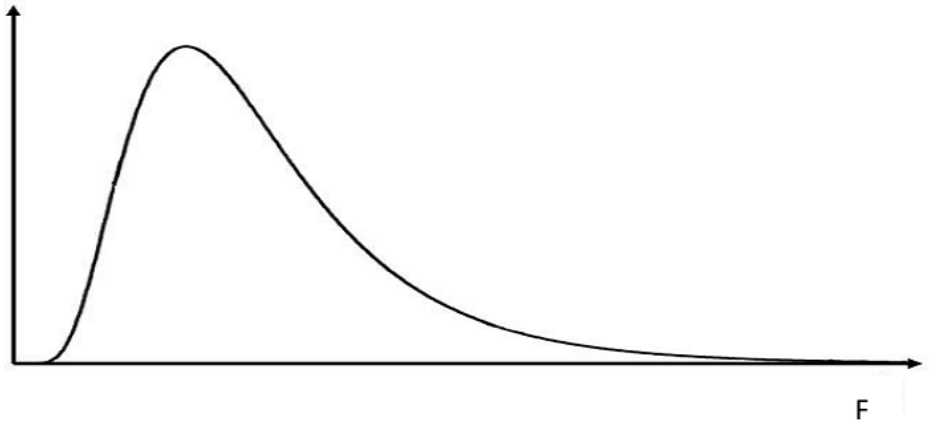
$\boldsymbol{v_1}$: degrees of freedom in the numerator $\boldsymbol{v_1} = \boldsymbol{n_1}-1$ // $\boldsymbol{n_1}$ First sample size .

$\boldsymbol{v_2}$ درجة حرية المقام $\boldsymbol{v_2} = \boldsymbol{n_2}-1$ // $\boldsymbol{n_2}$ حجم العينة الثانية

$\boldsymbol{v_2}$: degrees of freedom in the station $\boldsymbol{v_2} = \boldsymbol{n_2}-1$ // $\boldsymbol{n_2}$ Second sample size .

كما يأخذ التمثيل البياني لكثافة توزيع فيشر شكلًا مشابهًا لتوزيع كاي تربيع، كما هو موضح أدناه

The graphical representation of the density of the Fisher distribution also takes a similar form to the chi-squared distribution, as shown below :



من الشكل البياني للتوزيع نستنتج ان التوزيع يتميز بالخصائص التالية :

From the graph of the distribution, we can conclude that the distribution is characterized by the following properties:

- توزيع غير متماثل حيث يميل الى اليمين ، ويكون دائما موجبا

-It is Asymmetrical distribution, tilting to the right always positive.

- يتأثر شكله بعدد درجات الحرية ، حيث إنه كلما زاد عدد درجات الحرية اقترب شكل توزيع فيشر من شكل التوزيع الطبيعي.

-It is shape is affected by the number of degrees of freedom, as the greater the number of degrees of freedom, the closer the shape of the Fisher distribution is to the shape of the normal distribution

2.2 . اختبار الفرضيات واتخاذ القرارات الإحصائية
2.2 Hypothesis testing and statistical decision-making

تعد صياغة الفرضيات في البحوث العلمية خطوة تأتي بعد تحديد الاشكالية، حيث تمثل الفرضية حلاً مؤقتاً يقترحه الباحث بشكل أولي، ليقوم لاحقاً بإثباتها أو دحضها من خلال عملية البرهنة والاختبار

Formulating hypotheses in scientific research is a step that comes after identifying the problem. A hypothesis represents a tentative solution initially proposed by the researcher, who will later prove or disprove it through the process of demonstration an testing .

الفرضية البحثية تعرّف على أنها فرضية تُبنى استناداً إلى أسس نظرية أو على ملاحظات وتجارب سابقة. لذلك، تتضمن صياغة إجرائية تهدف إلى اختبارها باستخدام الملاحظة العلمية المنظمة. على سبيل المثال: يمكن صياغة فرضية تفيد بأن الممارسة الرياضية تساهم في تحسين الصحة النفسية لأطفال المرحلة الابتدائية.

أما بالنسبة **لاختبار الفرضيات الإحصائية**، فهو يمثل خطوة أساسية للتحقق من صحة الفرضيات بناءً على البيانات والتحليل العلمي.

A research hypothesis is defined as a hypothesis built upon theoretical foundations or previous observations and experiments. Therefore, it involves a procedural formulation aimed at testing it using systematic scientific observation. For example, a

hypothesis might be formulated stating that physical activity contributes to improving the mental health of elementary school children.

As for **statistical hypothesis testing**, it represents a fundamental step in verifying the validity of hypotheses based on data and scientific analysis.

1.2.2 خطوات اختبار الفرضيات
2.2.1 Steps for testing hypotheses

من المنظور الإحصائي، تُعَدُّ الفرضيات تصورًا أو افتراضًا لعلاقات محتملة بين متغيرين أو أكثر .تشكّل هذه الفرضيات حلولًا مبدئية لقضايا أو مشكلات معينة يركّز عليها البحث، و يتم استمدادها من الإطار النظري للدراسة ومتطلبات تحليلها .ويتم التأكد من صحة هذه الفرضيات عن طريق اختبارها إحصائيًا، وتضم ستة خطوات :

From a statistical perspective, hypotheses are hypotheses that represent a conception or assumption of possible relationships between two or more variables. These hypotheses constitute preliminary solutions to specific issues or problems addressed by the research, and they are derived from the theoretical framework of the study and the requirements of its analysis. The validity of these hypotheses is confirmed through statistical testing, it includes six steps :

الخطوة الاولى : تحديد الفرضيات

Step 1:the Hypotheses of Defining

بشكل عام تنقسم الفرضيات الاحصائية الى نوعين:

Generally, statistical hypotheses are divided into types two

الفرض الصفرى: يشير هذا النوع من الفروض إلى غياب أي علاقة بين المتغيرات،أو انعدام الفرق بين المعالجات المتعلقة بمتغير معين . يعرف أيضًا بالفرض الأصلي لأنه يتضمن شرط عدم التغييرويرمز له بالرمز H_0 فمثلا "لا يوجد فرق في متوسط الرضا الوظيفي بين الذكور والإناث" أو" لاتوجد علاقة بين عدد ساعات الدراسة والتحصيل الاكاديمي "

The null hypothesis: This type of hypothesis indicates the absence of any relationship between variables, or the lack of any difference between treatments related to a specific variable. It is also known as the original hypothesis because it includes the condition of no change and is denoted by the symbol H_0. For example "There is no difference in average job satisfaction between males and females" or "There is no relationship between the number of There is nostudy hours and academic achievement."

الفرض البديل: يمثل الفرض البديل تفسيرًا بديلا للظاهرة قيد الدراسة ، حيث يعتبر فرضًا يقوم الباحث باختبار صحته .يهدف إلى دحض الفرض الصفري ويُقر بوجود علاقة بين متغيرين أو أكثر، أو وجود اختلاف بين المعالجات .ويرمز إليه بالرمز H_1 مثلا يوجدعلاقة بين الرضا بين الموظفين المتزوجين والعزاب

The alternative hypothesis: The alternative hypothesis represents an alternative explanation for the phenomenon under study. It is a hypothesis that the researcher tests. It aims to refute the null hypothesis and affirms the existence of a relationship between two or more variables, or a difference between treatments. It is denoted by the symbol H_1. For example, there is a relationship between job satisfaction between married and single employees.

الخطوة الثانية : تحديد الدلالة الإحصائية (مستوى الدلالة)

Step 2 : Defining of Statistical Significance (Level of Significance)

أثناء اختبار الفروض، يتم الحصول على قيم إحصائية تساعد الباحث على اتخاذ قرار برفض أو قبول الفرض الصفري .تُعد القيمة الدالة إحصائيًا هي تلك التي تتجاوز القيمة الحرجة .وفي العلوم الاجتماعية والإنسانية، تُستخدم بشكل شائع مستويان للدلالة الإحصائية :

During hypothesis testing, statistical values are obtained that help the researcher decide whether to reject or accept the null hypothesis. A statistically significant value is one that exceeds the critical value. In the social sciences and humanities, two levels of statistical significance are commonly used:

المستوى الأول 0.05 : أو 5%، حيث تشير النتيجة إلى أن 95%من القيم يمكن تفسيرها بناءً على عوامل تجريبية

Level 1: 0.05 or 5%, where the result indicates that 95% of the values can be explained based on experimental factors.

المستوى الثاني 0.01 : أو 1%، ويُستخدم هذا المستوى عندما يرغب الباحث في زيادة قوة الاختبار الإحصائي، مما يعني أن 99%من النتائج حقيقية وتعود لمعالجات تجريبية .

Level 2: 0.01 or 1%. This level is used when the researcher wants to increase the power of the statistical test, meaning that 99% of the results are valid and attributable to experimental treatments.

الخطوة الثالثة : اختيار الاختبار الإحصائي المناسب وحساب القيمة المحسوبة

Step 3: Choosing the appropriate statistical test and calculating the calculated value

يتم الاختيار بناءً على معايير خاصة كطبيعة الهدف البحثي (علاقة ، فرق ، ...)و خصائص المتغيرات (اسمي ، ترتيبي، ...)وخصائص العينة و طريقة السحب الخاصة.

بعد اختيار الاختبار المناسب، تأتي المرحلة الحسابية التي تحول البيانات الخام إلى قيمة رقمية ذات معنى إحصائي اي حساب **القيمة المحسوبة** وهي تنتج عن تطبيق الصيغة الرياضية للاختبار المختار على بيانات العينة . تمثل مقياساً للانحراف أو الاختلاف أو القوة الملاحظة في البيانات، مُقارنةً بما هو متوقع تحت الفرضية الصفرية.

The selection is based on specific criteria such as the nature of the research objective (relationship, difference, etc.), the characteristics of the variables (nominal, ordinal, etc.), the characteristics of the sample, and the specific sampling method.

After selecting the appropriate test, the next step is computation, which transforms the raw data into a statistically meaningful numerical value. This is known as the **calculated value**, which results from applying the mathematical formula of the chosen test to the sample data. It represents a measure of the observed deviation, difference, or strength in the data, compared to what is expected under the null hypothesis.

الخطوة الرابعة : تحديد اتجاه الفرضية

Step 4: Determining the direction of the hypothesis

في اختبار الفروض الإحصائية، يتم تحديد منطقتين رئيسيتين بناءً على النتائج الإحصائية :

In statistical hypothesis testing, two main zones are identified based on the statistical results:

منطقة القبول :وهي المسافة التي تحتفظ بالقيم التي تقرر قبول الفرض الصفري فيها، ويتم تحديدها بقيمة حرجة تفصلها عن المنطقة الأخرى .

The Acceptance Zone: This is the range of values that determines the acceptance of the null hypothesis. It is defined by a critical value separating it from the other zone.

منطقة الرفض :تُعرف أيضًا باسم المنطقة الحرجة، وتحتوي على القيم الإحصائية التي تدفع الباحث إلى رفض الفرض الصفري وقبول الفرض البديل.

The Rejection Zone: Also known as the critical zone, this contains the statistical values that lead the researcher to reject the null hypothesis and accept the alternative hypothesis.

الفرضية الصفرية لا يمكن تقييمها ، فهي نافية .لكن بالنسبة للفرضية البديلة فلها ثلاث اتجاهات على النحو التالي

The null hypothesis cannot be evaluated; it is negative. However, the alternative hypothesis has three directions, as follows

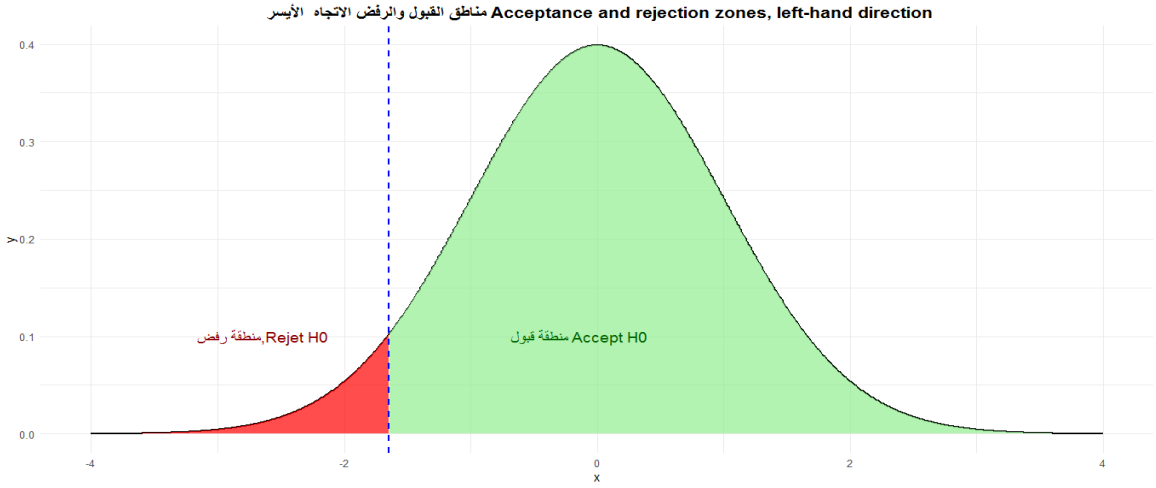
أ.الاختبار الأحادي (الطرف الموجه)

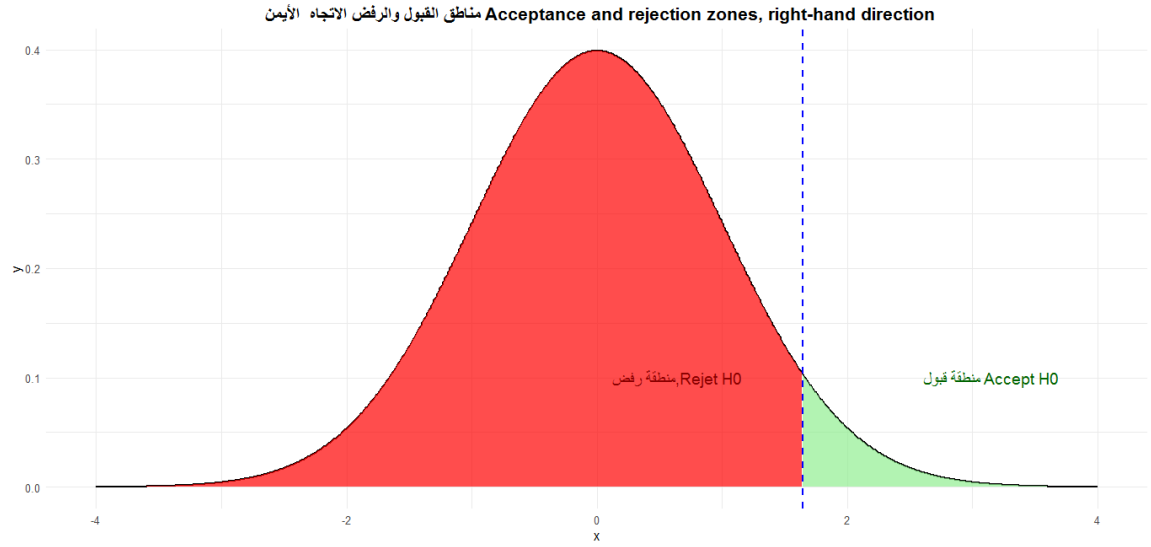
A.One-sided (directional) test

يُستخدم هذا النوع من الاختبارات لتحديد اتجاه الفروق بين العينات في الدراسة .في هذا الاختبار، يركز الباحث فقط على طرف واحد من توزيع القيم بدلاً من الأطراف المتعددة، بحيث تقع نسبة من القيم في طرف واحد، قد يكون متجه نحو اليمين او متجه نحو اليسار كما هو موضح في المنحنى التوضيحي.

This type of test is used to determine the direction of differences between samples in a study. In this test, the researcher focuses only on one end of the instead of multiple ends, distribution of values

such that a proportion of the values fall at one end. It may be facing right or facing leftas shown in the illustrative curve.





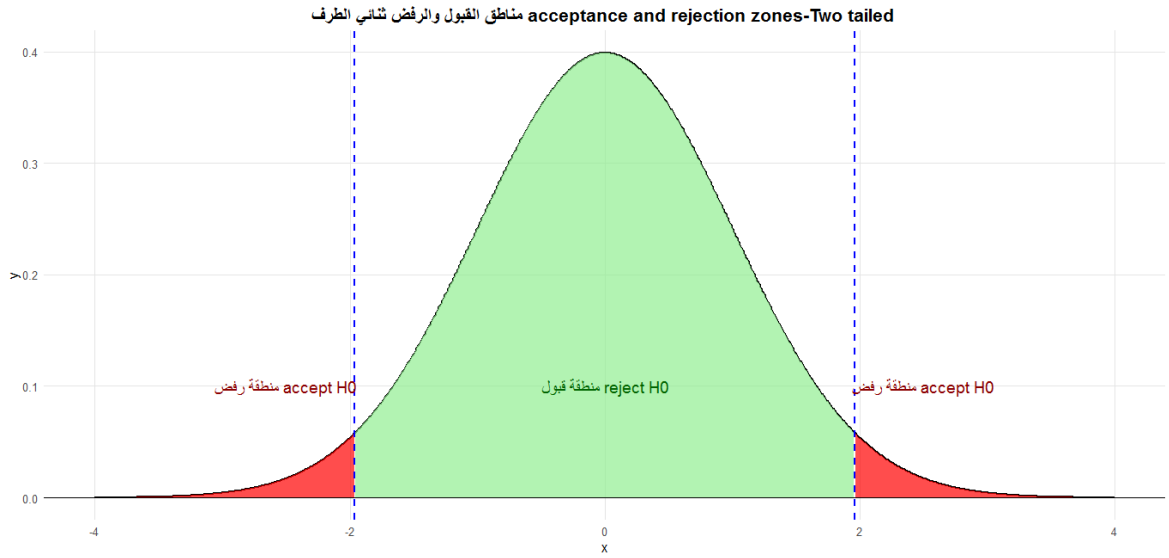
ب. الاختبار الثنائي الطرف (غير موجه)

B. Two-tailed (non-directional) test

يعتمد هذا الاختبار على مقارنة الفروق بين مجموعتين أو أكثر دون صياغة توقع مسبق حول اتجاه هذه الفروق. على سبيل المثال، عندما يفترض الباحث أن متوسط قيم مجموعتين

" تجريبية وضابطة " غير متساوٍ ويمكن أن يميل لأي جانب، يُطلق على هذا الاختبار غير الموجه .يتميز بوجود منطقتين حرجتين في الطرف الأيمن والأيسر لمنحنى التوزيع التماثلي.كما هو موضح في المنحنى التوضيحي.

This test compares the differences between two or more groups without making any prior predictions about the direction of these differences. For example, when a researcher assumes that the mean are " experimental and control" of two groups values unequal and could be tilted in either direction, this is called a non-directional test. It is characterized by the presence of two critical regions at the right and left ends of the symmetrical distribution curve, as illustrated in the diagram.



الخطوة الخامسة: اتخاذ القرار الإحصائي

Step 5: Make the Statistical Decision

بعد حساب القيمة المحسوبة نقوم بالمقارنة بينها وبين

القيمة الحرجة اذا كانت القيمة المحسوبة اكبر من القيمة الحرجة نرفض H_0 واذا كانت القيمة المحسوبة اقل من القيمة الحرجة نقبل H_0

After calculating the calculated value, we compare it to the critical value. If the calculated value is greater than the critical value, we reject H_0. If the calculated value is less than the critical value, we accept H_0.

الخطوة السادسة : التفسير

Step 6 : Interpretation

تفسير النتيجة في سياق البحث اي عند رفض H_0 "توجد أدلة إحصائية كافية عند مستوى دلالة α"

عند عدم رفض H_0 " لا توجد أدلة إحصائية كافية عند مستوى دلالة α "

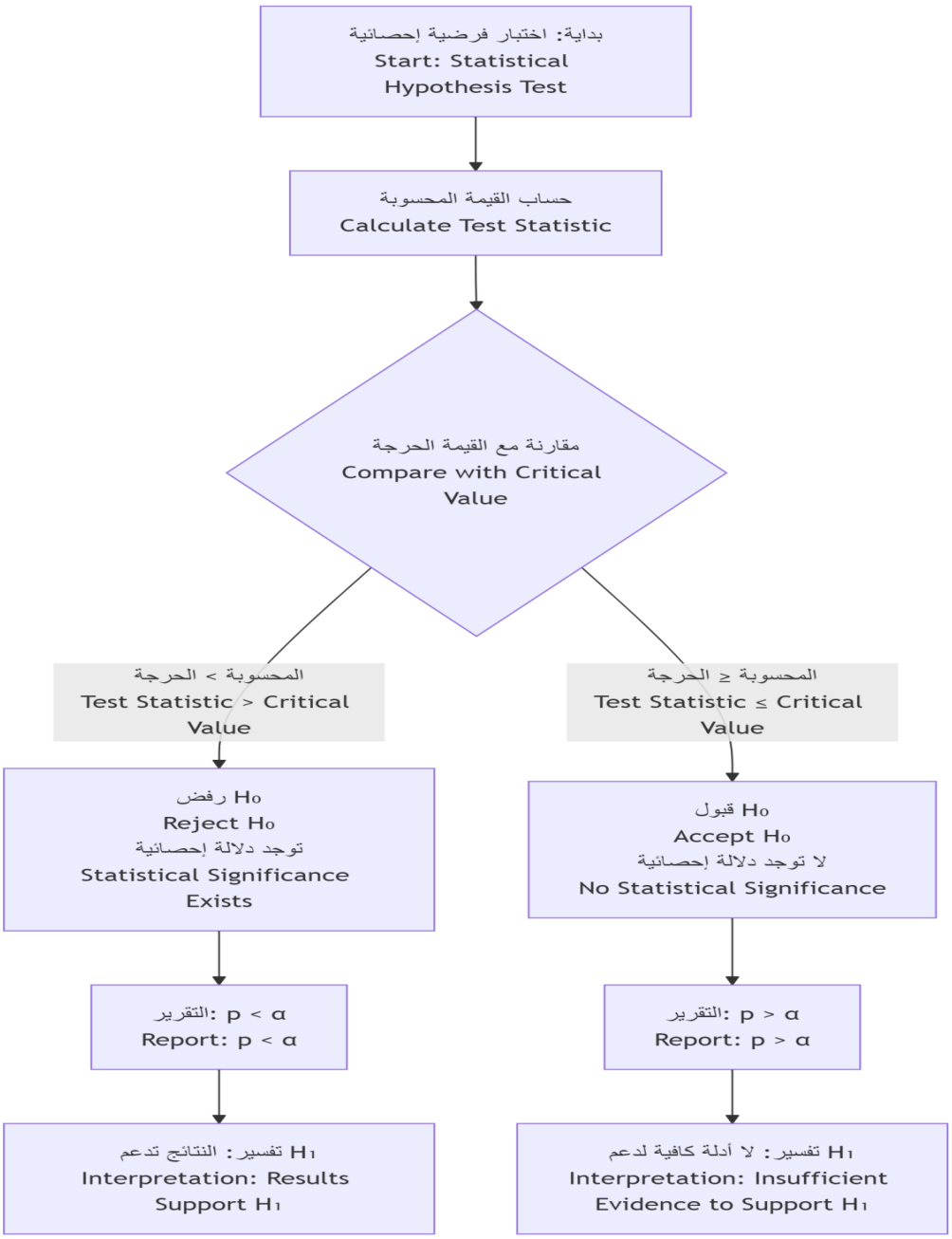
ملاحظة "عدم رفض H_0 " لا يعني أنها صحيحة، بل يعني أن البيانات الحالية لا توفر أدلة كافية ضدها. قد يكون السبب عينة صغيرة أو تباين كبير

Interpreting the result in the context of the research:

When H_0 is rejected, "there is sufficient statistical evidence at the significance level of α." When $H_0$is not rejected, "there is insufficient statistical evidence at the significance level of α. Note that "H_0 is not rejected" does not mean it is true, but rather that the current data do not provide sufficient evidence against it. This could be due to a small sample size or large variance.

يمكن تلخيص خطوات الاختبار الاحصائي في المخطط التالي

The steps of the statistical test can be summarized in the following diagram


بداية: اختبار فرضية إحصائية
Start: Statistical Hypothesis Test
حساب القيمة المحسوبة
Calculate Test Statistic
مقارنة مع القيمة الحرجة
Compare with Critical Value
المحسوبة > الحرجة
Test Statistic > Critical Value
المحسوبة ≤ الحرجة
Test Statistic ≤ Critical Value
رفض H₀
Reject H₀
توجد دلالة إحصائية
Statistical Significance Exists
قبول H₀
Accept H₀
لا توجد دلالة إحصائية
No Statistical Significance
التقرير: p < α
Report: p < α
التقرير: p > α
Report: p > α
تفسير: النتائج تدعم H₁
Interpretation: Results Support H₁
تفسير: لا أدلة كافية لدعم H₁
Interpretation: Insufficient Evidence to Support H₁

.

2.2المعايير المنهجية الحاكمة لاختيار الاختبار الإحصائي

2.2 Methodological Criteria Governing the Selection of Statistical Test

اختيار الاختبار الإحصائي الملائم يعتمد على ثلاثة عوامل أساسية مترابطة تشكل حجر الأساس لاتخاذ القرارات الإحصائية الصحيحة. استخدام اختبار غير ملائم يمكن أن يؤدي إلى استنتاجات خاطئة، حتى إذا كانت البيانات ذات جودة عالية.

Choosing the appropriate statistical test depends on three fundamental, interconnected factors that form the cornerstone of sound statistical decision-making. Using an inappropriate test can lead to erroneous conclusions, even if the data is of high quality

1. فرضيات البحث 1 .Research Hypotheses

- **الفرضية البحثية:** تركز على تحديد طبيعة العلاقة التي يُراد دراستها، سواءً كانت وصفية، مقارنة، ارتباطية، أو تنبؤية.

-Research Hypothesis: Focuses on defining the nature of the relationship to be studied, whether descriptive, comparative, correlational, or predictive

- **الفرضيات الإحصائية:** الفرضية الصفرية (H_0)، الفرضية البديلة (H_1)

-Statistical Hypotheses: Null Hypothesis (H_0), Alternative Hypothesis (H_1)

-عدد المتغيرات: أحادية، ثنائية، متعددة

-Number of Variables: Univariate, Bivariate, Multiple

- **طبيعة الاختبار:** اختبارات معلمية (بارامترية) لا معلمية (لابارامترية)

-Type of test: Parametric Non-parametric

2. خصائص العينة والبيانات

2.Sample and Data Characteristics

- **حجم العينة:** عينات صغيرة ($n < 30$) غالباً تستخدم اختبارات لا معلمية اما العينات الكبيرة : تفضل الاختبارات المعلمية

-Sample Size: Small samples often use non-parametric tests, while large samples prefer parametric tests

- **نوع العينات:** مستقلة أم مرتبطة (مقترنة)

-Sample Type: Independent or paired

- **توزيع البيانات:** طبيعي أم غير طبيعي

-Data Distribution: Normal or non-normal

3. مستويات قياس البيانات

3. Levels of Data Measurement

- **البيانات الاسمية:** تصنيف بدون ترتيب (جنس، لون، فئة)، الاختبارات المناسبة هي كاي مربع، فيشر

-Nominal Data: Classification without order (gender, color, category). Suitable tests are chi-square and Fisher's rank correlation coefficient

- **البيانات الترتيبية:** ترتيب بدون مسافات متساوية (مستويات الرضا، التصنيفات)، الاختبارات المناسبة هي سبيرمان

-Ordinal Data: Ordering without equal spacing satisfaction levels, ratings). Suitable tests are) Spearman's rank correlation coefficient

-البيانات الفترية/النسبية: مسافات متساوية، نقطة صفر حقيقية (درجات الحرارة، الطول، الوزن)، الاختبارات المناسبة هي بيرسون، الانحدار

-Interval/Relative Data: Equal spacing, true zero point (temperature, height, weight). Suitable tests are Pearson's rank correlation coefficient and regression

في الاخير يتم اتخاذ القرار لاختيار الاختبار الإحصائي بناءا على

Finally, the decision to choose a statistical test is based on

1 فحص العلاقات والارتباطات:

1 Examining Relationships and Correlations.

متغيران كميان (فترى/نسبي): اذا كانت علاقة خطية نلجأ الى معامل ارتباط بيرسون أما علاقة رتيبة: معامل ارتباط سبيرمان

Two quantitative variables (interval/relative) :If the relationship is linear, we use Pearson's correlation coefficient. If the relationship is monotonic, we use Spearman's rank correlation coefficient

متغيران فئويان: اختبار كاي مربع، معامل فاي

Two categorical variables: Chi-square test and phi coefficient

أكثر من متغيرين: الانحدار المتعدد

More than two variables: Multiple regression

2. التنبؤ: 2.Prediction :

الانحدار الخطي البسيط والمتعدد

Simple and Multiple Linear Regression

تمارين Exercices

التمرين الأول **Exercice 1**

تخضع أوزان عبوات إحدى الأنواع من الطماطم التوزيع الطبيعي وسطه 65 وانحرافه المعياري 15 غ

The weights of jars of a certain type of tomato are subject to a normal distribution with a mean of 65 and a standard deviation of 15 g.

ما هو احتمال أن وزن إحدى العبوات الذي أخذ عشوائيا يزيد عن 55 غ ؟

What is the probability that the weight of a randomly selected jar is greater than 55 g?

ما هو احتمال أن وزن إحدى العبوات الذي أخذ عشوائيا يقل عن 60 غ ؟

What is the probability that the weight of a randomly selected jar is less than 60 g?

التمرين الثاني **Exercice2**

إذا علمت أن الإنفاق الشهري لأفراد مجتمع ما يتوزع طبيعيا بمتوسط 75.49 دينار وبانحراف
معياري يساوي 82.11 دينار فإذا تم اختيار عينة عشوائية من 15 أفراد فما احتمال :

If the average monthly expenditure of individuals in a community is normally distributed at 75.49 dinars with a standard deviation of 82.11 dinars, and a

random sample of 15 individuals is selected, what is the probability that:

-أن يكون متوسط الإنفاق الشهري ما بين 75.47 و 75.51 ؟

-The average monthly expenditure will be between 75.47 and 75.51?

-أن يكون متوسط الإنفاق الشهري ما بين 89.44 و 89.54 ؟

-The average monthly expenditure will be between 89.44 and 89.54

التمرين الثالث **Exercice3**

بناءً على المعايير الممنهجة لاختيار الاختبار الإحصائي المناسب، صف بإيجاز كيف تختار الاختبار الإحصائي المناسب في الحالتين التاليتين

Based on the systematic criteria for selecting an appropriate statistical test, briefly describe how you would choose the appropriate statistical test in the following two cases:

الحالة الاولى : دراسة تهدف إلى مقارنة متوسط درجات الذكور والإناث في اختبار الذكاء

Case 1: A study aimed at comparing the average scores of males and females on an intelligence test.

الحالة الثانية : دراسة تهدف إلى معرفة ما إذا كان هناك ارتباط بين المستوى التعليمي للوالدين واتجاهات الابناء نحو التعلم عبر الإنترنت

Case 2: A study aimed at determining whether there is a correlation between the parents' educational level and their children's attitudes toward online learning.

التمرين الرابع **Exercice 4**

في دارسة احصائية حول تأثير برنامج تدريبي جديد على إنتاجية الموظفين ل 50 موظفًا. قمنا بقياس إنتاجيتهم قبل وبعد البرنامج

In a statistical study investigating the impact of a new training program on employee productivity, we measured the productivity of 50 employees before and after the program.

-اكتب الفرضية الصفرية والبديلة

- Write the null and alternative hypotheses

- وضح باختصار الفرق بين :

- Briefly explain the difference between

الاختبار أحادي الذيل (One-tailed test)

الاختبار ثنائي الذيل (Two-tailed test)

Chapter 3 الفصل الثالث

معاملات الارتباط

Correlation Coefficients

تمهيد **Preface**

يستخدم معاملات الارتباط في تقدير قوة العلاقة بين متغيرين أو أكثر وهي قيمة حقيقية خالية من وحدات قياس المتغيرات المرتبطة بالعلاقة وهناك عدة أنواع من الارتباط يختص كل منها بنوع معين من العلاقات، ويقتصر هذا الفصل على معاملي ارتباط للمتغيرات الكمية، و معاملات الخاص بالمتغيرات الوصفية

The correlation coefficients are used to estimate the strength of the relationship between two or more variables. It is a real value free of units of measurement of the variables associated with the relationship. There are several types of correlation, each of which is specific to a specific type of relationship. In this chapter will be limited to two correlation coefficients for quantitative variables, and a correlation coefficient for descriptive variables.

3-1 اساسيات معاملات الارتباط

3-1 Fundamentals of correlation coefficients

تعريف الارتباط **Definition of correlation**

هو تعيين طبيعة وقوة العلاقة بين متغيرين أو عدمها

It is to determine the nature and strength of the relationship between two variables or the lack there of.

تعريف معامل الارتباط

Definition of correlation coefficient

هو مؤشر هذه العلاقة

It is an indicator of this relationship.

تعريف المتغير المستقل والمتغير التابع

Definition of independent and dependent variables

المتغير المستقل: يُعرف بأنه المتغير الذي يمكن التحكم فيه وضبطه، ويتميز بكونه العنصر الفاعل، حيث قد يؤثر أو لا يؤثر على المتغير التابع .

Independent variable: This is defined as the variable that can be controlled and adjusted. It is the active element, as it may or may not influence the dependent variable.

المتغير التابع: يُقصد به المتغير الذي لا يمكن التحكم فيه، ويتصف بكونه العنصر المتأثر، حيث من المحتمل أن يتأثر أو لا يتأثر بالمتغير المستقل

Depende of variable This refers to the variable that cannot be controlled. It is the element that is influenced, as it may or may not be affected by the independent variable.

قياس معامل الارتباط

Correlation Coefficient Measurement

يتم قياس الارتباط بناءً على مستوى الاقتران بين بعدين رئيسيين ضمن التوزيع الطبيعي المعياري. يتراوح معامل الارتباط بين 1 - مروراً بـ 0 وصولاً إلى 1+. ومع ذلك، من النادر جداً في

مجالات العلوم الإنسانية والاجتماعية أن يتم الوصول إلى القيمة الكاملة +1. ويتم تحديد نوعية هذا الارتباط وفقاً للجدول التالي

Correlation is measured based on the level of correlation between two principal dimensions within a standard normal distribution. The correlation coefficient ranges from -1 through 0 up to +1 .However, it is very rare in the humanities and social sciences to reach the full value of +1. The quality of this correlation is determined according to the following table:

Type /النوع	Coefficient /المعامل	Interpretation /التفسير
ارتباط طردي تام Perfect Positive Correlation	+1	تغيير كامل ومتناسب في نفس الاتجاه Complete proportional change in the same direction
ارتباط طردي قوي Strong Positive Correlation	+0.70 إلى +0.99+ +0.70 to +0.99	علاقة قوية في نفس الاتجاه Strong relationship in the same direction
ارتباط طردي متوسط Moderate Positive Correlation	+0.40 إلى +0.69+ +0.40 to +0.69	علاقة متوسطة القوة في نفس الاتجاه Moderate strength relationship in the same direction
ارتباط طردي ضعيف Weak Positive Correlation	+0.10 إلى +0.39+ +0.10 to +0.39	علاقة ضعيفة في نفس الاتجاه Weak relationship in the same direction
شبه منعدم/ارتباط ضعيف جداً VeryWeak/Negligible Correlation	+0.01 إلى +0.09+ +0.01 to +0.09	علاقة ضعيفة جداً قد لا تكون ذات معنى عملي Very weak relationship, may not be practically meaningful
ارتباط منعدم No Correlation	0	لا توجد علاقة خطية No linear relationship exists
ارتباط عكسي ضعيف جداً Very Weak Negative Correlation	-0.01 إلى -0.09- -0.01 to -0.09	علاقة عكسية ضعيفة جداً Very weak inverse relationship
ارتباط عكسي ضعيف Weak Negative Correlation	-0.10 إلى -0.39- -0.10 to -0.39	علاقة عكسية ضعيفة Weak inverse relationship

النوع/ Type	المعامل/ Coefficient	التفسير/ Interpretation
ارتباط عكسي متوسط Moderate Negative Correlation	-0.69إلى -0.40 -0.40 to -0.69	علاقة عكسية متوسطة القوة Moderate strength inverse relationship
ارتباط عكسي قوي Strong Negative Correlation	-0.99إلى -0.70 -0.70 to -0.99	علاقة عكسية قوية Strong inverse relationship
ارتباط عكسي تام Perfect Negative Correlation	-1	تغيير كامل ومتناسب في اتجاهين متعاكسين Complete proportional change in opposite directions

3-2معاملات الارتباط للمتغيرات الكمية

3-2 Correlation coefficients for quantitative variables

يتناول دراسة العلاقة بين الظواهر القابلة للقياس الكمي أو العددي، بما يشمل جميع الظواهر التي يمكن التعبير عنها بشكل رقمي مثل كمية الإنتاج أو المساحة وغيرها. وينقسم هذا المجال إلى عدة أنواع:

This field studies the relationship between phenomena that can be quantified or numerically measured, including all phenomena that can be expressed numerically, such as production volume or area. It is divided into several types:

3-2-1 معامل الارتباط بيرسون

3-2-1 Pearson correlation coefficient

يتم استخدام معامل الارتباط بيرسون عندما يحتوي كلا المتغيرين، المستقل والتابع، على قيم حقيقية متصلة، بمعنى أن يكونا بمقياس نسبي وليس بقيم منفصلة أو رتبية أو تمثل فئات أو تكرارات. يُعد معامل الارتباط بيرسون واحداً من أبرز وأكثر المعاملات استخداماً في المجالات العلمية، خاصة في العلوم الإنسانية والاجتماعية، ويتم حسابه باستخدام المعادلة التالية:

Pearson's correlation coefficient is used when both variables, independent and dependent, have continuous, true values, that is, values on a relative scale, not discrete,

ordinal, or representing categories or frequencies. Pearson's correlation coefficient is one of the most prominent and widely used coefficients in scientific fields, particularly in the humanities and social sciences, and is calculated using the following equation:

$$r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

Where حيث

n: مجموع أفراد العينة

n: Total number of individuals in the sample

$\sum xy$ Sum of the product of x and y	$\sum xy$ مجموع حاصل ضرب x في y
$\sum x$ Sum of the variable X	$\sum x$ مجموع قيم المتغير X
$\sum y$ Sum of the variable Y	$\sum y$ مجموع قيم المتغير Y
$\sum x^2$ Sum of squares of the values x	$\sum x^2$ مجموع مربعات قيم المتغير x
$\sum y^2$ Sum of squares of the values y	$\sum y^2$ مجموع مربعات قيم المتغير y

الدلالة الإحصائية Statistical signification

لتحديد الدلالة الاحصائية بين المتغيرين، يتم تحديد اتجاه فرضية البحث أولاً، ثم حساب درجات الحرية وهي **N-2** سواءً لاختبار ذو الطرف الواحد أو الطرفين. يتم مقارنة النتيجة α بعد ذلك، يتم اختيار مستوى الدلالة المعنوية التجريبية مع النتيجة الجدولية بناءً على هذه الإعدادات.

إذا كانت النتيجة التجريبية مساوية أو أكبر من النتيجة الجدولية، يتم قبول الفرض البديل ورفض الفرض الصفري. بمعنى أن هناك علاقة ارتباطية بين المتغيرين. أما إذا كانت النتيجة التجريبية أقل من النتيجة الجدولية، يتم قبول الفرض الصفري ورفض الفرض البديل، مما يعني عدم وجود علاقة ارتباطية ذات دلالة إحصائية بين المتغيرين.

To determine the statistical significance between the two variables, the direction of the research hypothesis is first defined, and then the degrees of freedom are calculated **N-2** Next, the significance level **α** is chosen for either a one-way or two-way test. The experimental result is then compared with the tabulated result based on these settings. If the experimental result is equal to or greater than the tabulated result, the alternative hypothesis is accepted and the null hypothesis is rejected, indicating a correlation between the two variables. Conversely, if the experimental result is less than the tabulated result, the null hypothesis is accepted and the alternative hypothesis is rejected, meaning there is no statistically significant correlation between the two variables.

Example **مثال**

أراد باحث أن يدرس العلاقة الارتباطية بين الاكتئاب و تقدير الذات ل 10 من الطلاب ، ثم قام بتلخيص البيانات و تبويبها في الجول التالي :

A researcher wanted to study the correlation between depression and self-esteem in a 10 of students, then summarized and categorized the data in the following table:

لاكتئاب depression	1	8	2	1	12	5	4	10	2	3
تقدير الذات self-esteem	5	6	2	5	9	5	6	3	1	10

المطلوب **Question**

هل يوجد ارتباط بين الاكتاب و تقدير الذات عند مستوى الدلالة 0.05 ؟

Is there a correlation between the depression and self-esteem at the significance level 0.05 ?

اولا، نقوم بصياغة اشكالية البحث

- First, we formulate the research problem

1 صياغة الفرضيات **1Formulating hypotheses**

H_0 لا يوجد ارتباط بين الاكتاب و تقدير الذات

H_0 ;Nothing correlation between the depression and self-esteem

H_1 يوجد ارتباط بين الاكتاب و تقدير الذات

H_1 ;Exiting correlation between for the volume of production and volume of oil exports

2 الطريقة الاحصائية **2 - Statistical Method**

البيانات تشكل متغيرا كميا، لذا فإن اختبار الحالة هو ارتباط بيرسون

The data as form a quantitate variable, so, test situation is Pearson correlation

n	الاكتئاب(X) depression	تقدير الذات(Y) self-esteem	X^2	Y^2	XY
1	1	5	1	25	5
2	8	6	64	36	48
3	2	2	4	4	4
4	1	5	1	25	5
5	12	9	144	81	108
6	5	5	25	25	25
7	4	6	16	36	24
8	10	3	100	9	30
9	2	1	4	1	2
10	3	10	9	100	30
المجموع sum	$\Sigma X = 48$	$\Sigma Y = 52$	$\Sigma X^2 = 368$	$\Sigma Y^2 = 342$	$\Sigma XY = 281$

$$r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}$$

$$r_p = \frac{\mathbf{10}(\mathbf{281}) - (\mathbf{48})(\mathbf{52})}{\sqrt{\left(\mathbf{10}(368) - (\mathbf{48})^2\right)\left(\mathbf{10}(342) - (\mathbf{52})^2\right)}}$$

$$r_p = \frac{\mathbf{2810} - \mathbf{2496}}{\sqrt{(3680 - \mathbf{2304})(3420 - \mathbf{2704})}}$$

$$r_p = \frac{\mathbf{314}}{\sqrt{(1376)(\mathbf{716})}} = \frac{\mathbf{314}}{\sqrt{\mathbf{985216}}}$$

$$r_p \approx \frac{\mathbf{314}}{\mathbf{992.58}} \approx \mathbf{0.316}$$

هذا يشير إلى ارتباط طفيف موجب بين الاكتئاب وتقدير الذات في هذه العينة، أي أنه كلما زاد الاكتئاب زاد تقدير الذات قليلا

This suggests a slight positive correlation between depression and self-esteem in this sample; that is, the more depressed you are, the more your self-esteem improves slightly.

3- Statistical decision 3-القرار الاحصائي

البحث عن درجة الحرية

Finding the degrees of freedom

df=n-2=10-2=8

البحث عن القيمة الجدولية من الجدول الاحصائي بيرسون عند α=0.05 تساوي 2.30 (انظر الملحق 4)

Finding the tabulated value from the Pearson statistical table at α=0.05 equal 2.30 (see appendix 4)

$$r_p \textbf{ (tab) = 2.30} \quad \geq \quad r_p \textbf{ (cal) = 0.316}$$

نقبل الفرضية الصفرية Accept H_0	نرفض الفرضية البديلة Reject H_1
لا يوجد ارتباط بين الاكتئاب و تقدير الذات عند مستوى الدلالة **0.05**	Nothing correlation between depression and self-esteemat the significance level **0.05**

3-2-2 معامل الارتباط الجزئي

3-2-2 Partial correlation coefficient

يتم استخدام معامل الارتباط الجزئي عندما يكون لدى الباحث **ثلاثة متغيرات** تؤثر على خاصية معينة. يقوم في البداية بحساب معامل ارتباط بيرسون بين كل الباحث زوج من المتغيرات، ثم يسعى إلى تحديد العلاقة بين متغيري ن مع ضبط أو استبعاد تأثير المتغير الثالث. بمعنى آخر، يتيح معامل الارتباط الجزئي قياس الارتباط الحقيقي بين المتغيرين بعد إقصاء تأثير المتغير الثالث عليهما. يمكن حسابه باستخدام الصيغة التالية:

The partial correlation coefficient is used when a researcher has **three variables** affecting a particular characteristic. The researcher first calculates Pearson's correlation coefficient between each pair of variables, then seeks to determine the relationship between two variables while controlling for or excluding the influence of the third variable. In other words, the partial correlation coefficient allows for measuring the true correlation between two variables after eliminating the influence of the third variable. It can be calculated using the following formula:

$$r_{12,3} = \frac{r_{12} - r_{13} * r_{23}}{\sqrt{(1 - r_{13}^{\,2}) * (1 - r_{23}^{\,2})}}$$

where **حيث :**

r_{12} : معامل الارتباط بين المتغير الأول والثاني

r_{12} : Correlation coefficient between the first and second variables

r_{13} : معامل الارتباط بين المتغير الأول والثالث

r_{13} : Correlation coefficient between the first and third variables

r_{23} : معامل الارتباط بين المتغير الثاني والثالث

r_{23} : Correlation coefficient between the second and third variable

الدلالة الإحصائية Statistical signification

يمكن حساب الدلالة الاحصائية بالصيغة التالية :

Statistical significance can be calculated by following formula :

$$T=\frac{r}{\sqrt{\frac{1-r^2}{n-k}}}$$

ومقارنتها مع القيمة الجدولية عند درجة الحرية **(N-k)** ، حيث **k** عدد المتغيرات ومستوى الدلالة **(0.05-0.01)** من جدول ستودنت

And comparing with the tabular value at the degree of freedom **(N-k)**, where **k**: number of variable at signification on **(0.01 - 0.05)** for Student test table.

مثال Example

اراد باحث معرفة العلاقة بين القلق ،الاكتئاب والانفصام عند مستوى دلالة 0.05 علماً أن معامل الارتباط بين القلق و الاكتئاب **0.45** وعلاقة القلق والانفصام **0.35** وعلاقة الاكتئاب والانفصام **0.65** وذلك لعينة حجمها **100.**

A researcher wanted to determine the relationship and schizophrenia at a between anxiety, depression given that the **0.05,** significance level of correlation coefficient between anxiety and and the relationship between **0.45** depression was **0.35 ,** the anxiety and schizophrenia was

relationship between depression and schizophrenia was **0.65** for a sample size is **100**

Solution **الحل**

1 .علاقة القلق بالاكتئاب بعد استبعاد الانفصام

1.The relationship between anxiety and depression after ruling out schizophrenia

$$r_{12,3} = \frac{r_{12} - r_{13} * r_{23}}{\sqrt{(1 - r_{13}^{\,2}) * (1 - r_{23}^{\,2})}}$$

$$r_{12,3} = \frac{0.45 - 0.35 * 0.65}{\sqrt{(1 - 0.35^2) * (1 - 0.65^2)}}$$

$$r_{12,3} = \frac{0.45 - 0.2275}{\sqrt{(1 - 0.1225) * (1 - 0.4225)}} = \frac{0.2225}{\sqrt{(0.8775)(0.5775)}}$$

$$r_{12,3} = \frac{0.2225}{0.7119} \approx 0.3126$$

2 .علاقة القلق بالانفصام بعد استبعاد الاكتئاب

2.The relationship between anxiety and schizophrenia after ruling out depression

$$r_{13,2} = \frac{r_{13} - r_{12} * r_{23}}{\sqrt{(1 - r_{12}^{\,2}) * (1 - r_{23}^{\,2})}}$$

$$r_{13,2} = \frac{0.65 - 0.45 * 0.35}{\sqrt{(1 - 0.45^2) * (1 - 0.35^2)}} = \frac{0.4925}{0.8365}$$

$$r_{13,2} \approx 0.58890$$

3 .علاقة الانفصام بالاكتئاب بعد استبعاد القلق

3.The relationship between schizophrenia and depression after ruling out anxiety

$$r_{23,1} = \frac{r_{23} - r_{12} * r_{13}}{\sqrt{(1 - r_{12}{}^2) * (1 - r_{13}{}^2)}}$$

$$= \frac{0.35 - 0.45 * 0.65}{\sqrt{(1 - 0.45^2) * (1 - 0.65^2)}}$$

$$r_{23,1} \approx 0.085$$

ويتم اختبار الدلالة الاحصائية لمعامل الارتباط الجزئي مثلا $r_{12,3}$ بالاختبار الثاني وفق الصيغة الاتية

The statistical significance of the partial correlation coefficient, for example, $r_{12,3}$ is tested using the second test according to the following formula

$$T = \frac{r}{\sqrt{\frac{1 - r^2}{n - k}}} = \frac{0.3126}{\sqrt{\frac{1 - 0.3126^2}{100 - 3}}} = \frac{0.3126}{\sqrt{\frac{1 - 0.3126^2}{100 - 3}}}$$

$$T = \frac{0.3126}{\sqrt{\frac{1 - 0.0977}{97}}} = \frac{0.3126}{\sqrt{\frac{0.9023}{97}}}$$

$$T = \frac{0.3126}{0.0964} \approx 3.24$$

بعدها نقارنها بالقيمة الجدولية عند مستوى دلاله **0.05** ودرجة حرية **97** وهي **1.98**

ومنه نستنتج انه لا توجد علاقة بين القلق والاكتاب بعد استبعاد الانفصام

We then compare this to the tabulated value at a degrees of **97** and **0.05** significance level of From this,we conclude that **1.98.** which is, freedom

and there is no relationship between anxiety depression after excluding schizophrenia

3-3 معاملات الارتباط للمتغيرات النوعية

3-3 Correlation coefficients for qualitative variables

المتغيرات النوعية تشير إلى تلك المتغيرات التي لا يمكن قياسها بطريقة كمية، وإنما تظهر على شكل خصائص أو تصنيفات أو رتب. من الأمثلة على ذلك: الحالة الصحية للأفراد، الذوق، أو التفضيلات. وعلى الرغم من أن هذه المتغيرات الوصفية قد تؤثر على الظواهر الاقتصادية، إلا أنه يصعب قياس درجة ارتباطها باستخدام معامل الارتباط الخطي البسيط. لذلك ظهرت الحاجة إلى تطوير مقاييس جديدة تساهم في قياس العلاقة بين هذه المتغيرات النوعية. وفي هذا الإطار، توجد عدة مقاييس مُعتمدة لهذا الغرض.

Qualitative variables refer to those variables that cannot be measured quantitatively but rather appear as characteristics, classifications, or rankings. Examples include individuals' health status, tastes, or preferences. Although these descriptive variables may influence economic phenomena, it is difficult to measure their correlation using simple linear correlation coefficients. Therefore, the need arose to develop new measures that contribute to measuring the relationship between these qualitative variables. Several approved measures exist for this purpose.

3-3-1 معامل الارتباط الرتب (سبيرمان)

3-3-1 Rank correlation coefficient)Spearman's)

تستند فكرة معامل الارتباط الرتبي إلى ترتيب مفردات كل متغير من المتغيرات الوصفية المدروسة إما بترتيب تصاعدي أو تنازلي. ويتم تخصيص قيمة رقمية لكل مفردة تعبر عن ترتيبها، وفق العلاقة التالية :

The ordinal correlation coefficient is based on ranking the items of each of the studied descriptive variables in either ascending or descending order. A numerical value is

assigned to each item to represent its ranking, according to the following formula:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

where حيث :

D : الفرق بين رتب نفس الفرد على المتغير الأول والثاني

D : The difference between the ranks of the same individual on the first and second variables.

Statistical signification الدلالة الإحصائية

يمكن حساب الدلالة الاحصائية بالصيغة التالية :

Statistical significance can be calculated by following formula :

$$T = r_s \sqrt{\frac{n}{1-r_s^2}}$$

ومقارنتها مع القيمة الجدولية عند درجة الحرية **N** ، ومستوى الدلالة **α** من جدول سبيرمان

And comparing with the tabular value at the degree of freedom **N**, and signification **α** for Student test table.

Example مثال

أوجد معامل ارتباط الرتب لتقديرات الطلاب في كل من مادتي علم الاجتماع وعلم النفس

Find the rank correlation coefficient for student grades in both sociology and psychology

علم الاجتماع (X) Sociology	A	C	B	A	C	C
علم النفس (Y) Psychology	A	C	C	B	C	A

الحل **Solution**

نرتب تقديرات الطلاب ترتيباً تصاعدياً أي بدءاً من التقدير A وفي حال تساوي تقديرين في الرتبة في مثالنا التقدير نقوم بقسمة الرتبتين على اثنين ونثبت المتوسط الحسابي أما بالنسبة ل C نقوم بقسمة الرتب الثلاث على ونثبت المتوسط أيضاً

We arrange the students' grades in ascending order, starting with grade A. If two grades have the same rank in our example, A, we divide the two ranks by two and keep the arithmetic mean. For C, we divide the three ranks by two and also keep the arithmetic mean

Student الطالب	**x**	**y**	**R(x) رُتبة x**	***R(Y)* رُتبة y**	**D=R(x)-R(Y)**	D^2
1	A	A	5.5	6.5	-1	1
2	C	C	2	2.5	-0.5	0.25
3	B	C	4	1.5	1.5	2.25
4	A	B	5.5		0.5	0.25
5	C	C	2	2.5	-0.5	0.25
6	C	A	2	6.5	-4.5	20.25
المجموع sum						**24.25**

$$r_s = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6(24.25)}{6(6^2-1)}$$

$$r_s = 1 - \frac{145.5}{210} \approx 0.307$$

يشير معامل سبيرمان الإيجابي الضعيف إلى وجود ارتباط طردي ضعيف بين تصنيفات الطلاب في علم الاجتماع وعلم النفس.

A weakly positive Spearman's coefficient indicates a weak positive correlation between student rankings in sociology and psychology

ثم نحول معامل الارتباط سبيرمان إلى مقياس T

من خلال المعادلة التالية

Then we convert Spearman's correlation coefficient to a T score using the following equation :

$$T = r_s \sqrt{\frac{n}{1-r_s^2}} = 0.307\sqrt{\frac{6}{1-0.307^2}} = 0.307\sqrt{\frac{6}{0.9057}} = 0.307(2.574)$$

$$T = 0.7902$$

ويتم اختبار الدلالة الاحصائية لمعامل الارتباط الرتبي سبيرمان عند مستوى الدلالة **0.01**

و درجة الحرية **6** في جدول سبيرمان نجد القيمة الجدولية هي **1** (انظر الملحق **5**) و هي اكبر

من 0.7902 القيمة المحسوبة و منه نستنتج انه لا يوجد ارتباط بين درجات الطلاب في علم الاجتماع وعلم النفس

The statistical significance of Spearman's rank correlation coefficient is tested at a significance level degrees of freedom. In Spearman's table, the 6 and 0.01 of which is greater than the calculated ,1 tabulated value is value of**0.7902** .(see Appendix 5).Therefore, we conclude

that there is no correlation between students' scores in sociology and psychology.

Remark : **ملاحظة :**

ان المقارنة بين معامل ارتباط الرتب ومعامل ارتباط بيرسون تظهر بعض الاختلافات البارزة. حيث معامل ارتباط سبيرمان يعتمد في حساباته على ترتيب البيانات بدلاً من استخدام القيم الفعلية، مما يعني أن تغير القيم لا يؤثر على قيمة معامل الارتباط. أما بالنسبة لمعامل ارتباط بيرسون، فإن أي تغيير في القيم يؤثر بشكل مباشر على قيمة معامل الارتباط الخاصة به.

Comparing rank correlation coefficients with Pearson's correlation coefficient reveals some key differences. Spearman's correlation coefficient relies on the ranking of data rather than the actual values, meaning that changes in values do not affect the correlation coefficient. In contrast, any change in values directly affects Pearson's correlation coefficient.

3-3-2 معامل الارتباط كاندل

3-3-2 Kendall correlation coefficient

معامل ارتباط كندل يُعد أداة لقياس درجة التناسق أو الاتفاق بين ترتيبين لزوج من المتغيرات الترتيبية. يُستعمل لتحديد مدى قوة العلاقة الرتيبة بينهما، دون الحاجة لافتراض وجود علاقة خطية أو نوع معين من التوزيع للبيانات، وفق العلاقة التالية

Kendall's correlation coefficient is a tool for measuring the degree of consistency or agreement between two arrangements of a pair of ordinal variables. It is used to determine the strength of the monotonic relationship between them, without needing to assume a linear relationship or a specific type of data distribution, according to the following fomula:

$$r_K = \frac{C - D}{C + D}$$

where **حيث :**

C: Sum compatibility ranks	**C : مجموع رتب التوافق**
D : Sum dissonance ranks	**D: مجموع رتب التنافر**

Calculation Steps **خطوات الحساب**

-نرتب البيانات بالنسبة للمتغير **x**

- We arrange the data with respect to the variable **x**

-حساب عدد الأزواج المتوافقة **C** وغير المتوافقة **D**

لكل زوج (i,j) حيث i<j

-Calculate the number of compatible pairs **C** and incompatible pairs **D**

متوافق : إذا كان ترتيب **X** و **Y** متناسقين

Concordant: If the order of **X** and **Y** is consistent

$$(x_i < x_j \text{ أو } \boldsymbol{or} \quad y_i < y_j); (x_i > x_j \text{ أو } \boldsymbol{or} \quad y_i > y_j)$$

غير متوافق : إذا كان ترتيب **X** و **Y** غير متناسقين

Discordant: If the order of **X** and **Y** is consistent

$$(x_i < x_j \text{ أو } \boldsymbol{or} \quad y_i > y_j); (x_i < x_j \text{ أو } \boldsymbol{or} \quad y_i > y_j)$$

A tie: if **تعادل : إذا كان**

$$(x_i = x_j \text{ أو } \boldsymbol{or} \quad y_i = y_j)$$

Statistical signification **الدلالة الإحصائية**

يمكن حساب الدلالة الاحصائية بالصيغة التالية :

Statistical significance can be calculated by following formula :

$$Z=\frac{r_k}{\frac{2(2n+5)}{9n(n-1)}}$$

والمقارنة على النحو التالي :

The comparison is as follows

(±1.96) at signification level 0.05	(±1.96) عند مستوى الدلالة 0.05
(±2.58) atsignification level 0.01	(±2.58) عند مستوى الدلالة 0.01

Exemple **مثال**

أجرى الباحث دراسة مشكلة الادمان على الانترنيت و علاقتها بالاكتئاب على مجموعة من الطلاب، وأسفرت النتائج على النحو التالي :

The researcher conducted a study on the problem of internet addiction and its relationship to depression on a group of students, and the results were as follows

الادمان Addiction X	**10**	**13**	**15**	**12**	**11**	**14**
الاكتئاب Depression Y	**63**	**57**	**54**	**33**	**64**	**50**

Required: **المطلوب**

- صغ إشكالية وفرضية بديلة لهذا البحث؟ استخدم الأسلوب الإحصائي لمعامل كاندل لهده الدراسة؟

Formulate a research problem and an alternative
Use the Kandel ? study hypothesis for this
coefficient as a statistical method for this study

Solution **الحل**

اشكالية البحث :هل توجد علاقة ارتباطية بين الإدمان على الانترنت والاكتئاب ؟

Research Problem: Is there a correlation between internet addiction and depression?

الفرضية البديلة :توجد علاقة ارتباطية دالة إحصائيا بين الإدمان على الانترنت والاكتئاب

Alternative hypothesis: There is a statistically significant correlation between internet addiction and depression.

الاسلوب الاحصائي هو معامل كاندل ويتم حسابه كالتالي

The statistical method

used is the Kendall coefficient, which is calculated as follow

$$r_k = \frac{C - \mathrm{D}}{C + \mathrm{D}}$$

اولا نرتب الأزواج حسب X تصاعديًا

First, We arrange the pairs according to **X** in ascending order.

(10,63), (11,64), (12,33), (13,57), (14,50), (15,54)

الجدول التالي يوضح كيفية حساب الأزواج المتوافقة C و الأزواج الغير المتوافقة D.

The following table shows how to calculate compatible C and incompatible pairs D .

(i, j)	المقارنة في X_i Compare in X_i	المقارنة في Y_i Compare in Y_i	النوع(C/D) Type (C/D)
(1,2)	$X_1 < X_2 \rightarrow 10 < 11$	$Y_1 < Y_2 \rightarrow 63 < 64$	C
(1,3)	$X_1 < X_3 \rightarrow 10 < 12$	$Y_1 > Y_3 \rightarrow 63 > 33$	D
(1,4)	$X_1 < X_4 \rightarrow 10 < 13$	$Y_1 > Y_4 \rightarrow 63 > 57$	D
(1,5)	$X_1 < X_5 \rightarrow 10 < 14$	$Y_1 > Y_5 \rightarrow 63 > 50$	D
(1,6)	$X_1 < X_6 \rightarrow 10 < 15$	$Y_1 > Y_6 \rightarrow 63 > 54$	D
(2,3)	$X_2 < X_3 \rightarrow 11 < 33$	$Y_2 > Y_3 \rightarrow 64 > 33$	D
(2,4)	$X_2 < X_4 \rightarrow 11 < 57$	$Y_2 > Y_4 \rightarrow 64 > 57$	D
(2,5)	$X_2 < X_5 \rightarrow 11 < 50$	$Y_2 > Y_5 \rightarrow 64 > 50$	D
(2,6)	$X_2 < X_6 \rightarrow 11 < 54$	$Y_2 > Y_6 \rightarrow 64 > 54$	D
(3,4)	$X_3 < X_4 \rightarrow 12 < 13$	$Y_3 < Y_4 \rightarrow 33 < 57$	C
(3,5)	$X_3 < X_5 \rightarrow 12 < 14$	$Y_3 < Y_5 \rightarrow 33 < 50$	C
(3,6)	$X_3 < X_6 \rightarrow 12 < 15$	$Y_3 < Y_6 \rightarrow 33 < 54$	C
(4,5)	$X_4 < X_5 \rightarrow 13 < 14$	$Y_4 > Y_5 \rightarrow 57 > 50$	D
(4,6)	$X_4 < X_6 \rightarrow 13 < 15$	$Y_4 > Y_6 \rightarrow 57 > 54$	D
(5,6)	$X_5 < X_6 \rightarrow 14 < 15$	$Y_5 < Y_6 \rightarrow 50 < 54$	C

نلاحظ ان عدد C هو 5 و عدد D هو 10 ،اذن

We observe that the number C is 5 and the number D is 10

therefore ,

$$r_K = \frac{5-10}{5+10} = \frac{-5}{15} = -0.333$$

المعامل يشير إلى **علاقة عكسية ضعيفة** بين الإدمان والاكتئاب في هذه العينة.

The coefficient indicates a weak inverse relationship between addiction and depression in this sample

نحسب الان الدلالة الاحصائية لمعامل الارتباط كاندل

We now calculate the statistical significance of Kandel's correlation coefficient

$$\mathbf{Z=\frac{r_k}{\frac{2(2n+5)}{9n(n-1)}} = \frac{-0.333}{\frac{2(2(6)+5)}{9(6)(6-1)}} = \frac{-0.333}{\frac{34}{270}} = \frac{-0.333}{0.126}}$$

$$\mathbf{Z \approx -2.643}$$

نقارن هذه القيمة المحسوبة مع القيمة الجدولية عند مستوى الدلالة **0.01** هي **2.58** و منه القيمة المحسوبة اقل من القيمة الجدولية وبالتالي نستنتج انه لا يوجد ارتباط بين ادمان الانتريت و الاكتئاب

We compare this calculated value with the tabulated value at a significance level

0.01 value , is **2.58** .

Therefore the calculated value is less than the , tabulated and we conclude that there is no correlation between internet addiction and depression.

Exercices تمارين

Exercice 1 **التمرين الاول**

في دراسة على مجموعة من الطلاب أجرى باحث عليهم اختبارين أحدهما يقيس القدرة على التحمل والثاني يقيس القدرة على التذكروحصل على النتائج الآتي

In a study conducted on a group of students, a researcher administered two tests: one measuring endurance and the other measuring memory ability. The following results were obtained.

القدرة على التحمل Endurance (x)	**24**	**12**	**5**	**18**	**32**	**17**	**10**
القدرة على التذكر Memory (y)	**21**	**15**	**23**	**13**	**7**	**17**	**23**

1 . احسب معامل ارتباط بيرسون

1. Calculate Pearson's correlation coefficient.

2. اختبر دلالة الارتباط بين المتغيرين عند مستوى الدلالة 0.05 وفسر النتيجة؟

2.Test the significance of the correlation between the two variables at the 0.05 significance level and interpret the result.

التمرين الثاني Exercice2

الجدول التالي يبين تقديرات 10 طلاب في امتحان علم الاجتماع والإحصاء

The following table shows the grades of 10 students in the sociology and statistics exam.

الطلاب Students	1	2	3	4	5	6	7	8	9	10
علم الاجتماع Social-ogy	جيد جدا Very good	مقبول Acce-ptabl	ممتاز Exce-llent	ممتاز Exce-llent	جيد جدا Very good	ممتاز Exce-llent	جيد good	جيد good	مقبول Accep-tabl	جيد جدا Very good
الاحصاء Statist-ics	ممتاز Exce-llent	ممتاز Exce-llent	جيد good	مقبول Acce-ptabl	جيد good	مقبول Acce-ptabl	مقبول Acce-ptabl	ممتاز Exce-llent	جيد good	جيد good

هل يوجد ارتباط بين تقديرات الطلاب عند مستوى الدلالة 0.01

Is there a correlation between student grades at the 0.01 significance level?

التمرين الثالث Exercice 3

اراد باحث معرفة العلاقة بين القلق (X) والاكتئاب (Y) والانفصام (Z)عند مستوى دلالة 0.05 لعينة عشوائية مكونة من 25 فرداً.و توصل إلى معامل الارتباط بين القلق والاكتئاب 0.30 و معامل الارتباط بين القلق والانفصام 0.45

ومعامل الارتباط بين الاكتئاب والانفصام 0.70

A researcher sought to determine the relationship between anxiety (X), depression (Y), and

schizophrenia (Z) at a significance level of 0.05 for a random sample of 25 individuals. The correlation coefficient between anxiety and depression was found to be 0.30, between anxiety and schizophrenia 0.45, and between depression and schizophrenia 0.70.

1.احسب معامل الارتباط الجزئي بين X و Y بعد ضبط تأثير المتغير Z

1. Calculate the partial correlation coefficient between X and Y after controlling for the effect of the variable Z .

2.ماذا تستنتج عن دور الانفصام (Z) في العلاقة بين القلق (X) والاكتئاب (Y)

2.What conclusions can you draw about the role of schizophrenia (Z) in the relationship between anxiety (X) and depression (Y)?

3.اختبر دلالة الارتباط الجزئي إحصائياً

3.Test the significance of the partial correlation statistically.

التمرين الرابع **Exercice4**

قام باحث بتقييم ترتيب 8 موظفين حسب الإنتاجية وحسب الروح المعنوية

A researcher evaluated the ranking of 8 employees according to productivity and morale.

الموظف Employee	A	B	C	D	E	F	I	G
الإنتاجية productivity	8	30	13	16	13	18	25	20
الروح المعنوية morale	7	10	22	11	9	19	13	14

1. احسب عدد الأزواج المتطابقة والأزواج غير المتطابقة

1.Calculate the number of concordant and the number of discordant.

2. احسب معامل كيندال واختبر الدلالة الاحصائية عند 0.05

2.Calculate Kendall's coefficient and test for statistical significance at 0.05.

Chapter 4 الفصل الرابع

تطبيقات اختبار كاي مربع (χ^2)

Applications of the Chi-2 (χ^2)

Preface تمهيد

يُعتبر اختبار كاي-مربع **(χ^2)** أداة أساسية في تحليل البيانات النوعية يجيب عن أسئلة جوهرية لا يمكن الإجابة عنها بالاختبارات المعلمية، مثل: "هل هناك علاقة بين التدخين وسرطان الرئة؟"

يُستخدم هذا الاختبار بشكل حصري مع البيانات التصنيفية أو البيانات التي تأتي على شكل تكرارات، مثل: عدد الأشخاص حسب فئة عمرية، عدد المنتجات حسب اللون، عدد الإجابات حسب الفئة (نعم/لا). لا يفترض توزيعاً طبيعياً للبيانات فهو اختبار إحصائي لا معلمي يستخدم لفحص الفروق بين التكرارات الملاحظة والتكرارات المتوقعة. يُستخدم في تطبيقين أساسيين:

The chi-square (χ^2) test is a fundamental tool in qual itative data analysis. It answers essential questions that cannot be answered by parametric tests, such as: "Is there a relationship between smoking and lung cancer?"

This test is used exclusively with categorical data such as the numbeo or data presented as frequencies, people by age group, the number of products by color, or the number of responses by category (yes/ no). It does not assume a normal distribution; it isa non-parametric statistical test used to examine differences between observed and predicted frequencies. It has two main applications:

اختبار حسن المطابقة لمقارنة توزيع عينة مع توزيع نظري مفترض واختبار الاستقلالية لفحص العلاقة بين متغيرين تصنيفين داخل جدول.

A goodness-of-fit test to compare a sample distributi on with a theoretically assumed distribution, and an independence test to examine the relationship between two categorical variables within a contingency table

4-1اختبار كاي-2 لحسن المطابقة

4-1The Chi-Two Test for Good Fit

يستخدم في تحديد ما إذا كان التوزيع الملاحظ لعينة ما يطابق (أو يختلف بشكل معنوي عن) توزيعاً نظرياً مفترضاً (مثل: توزيع منتظم، نسب معروفة مسبقاً، إلخ). يحسب بالصيغة التالية

It is used to determine whether the observed distribtion of a sample matches (or significantly differs from) a theoretically assumed distribution (e.g., a uniform distribution, predetermined proportions, etc.). It is calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

حيث **where**

O_i التكرار الملاحظ للفئة **i**	O_i :Observed frequency of class **i**
E_i: التكرار المتوقع	E_i: Expected frequency

يحسب بالصيغة التالية :

It is calculated using the following formula:

$$E_i = N * p_i$$

E_i=نسبة احتمال كل عنصر *العد الكلي

4.1.1 Its hypotheses 1.1.4فرضياته

الفرض الصفري (H_0): التوزيع الملاحظ يطابق التوزيع النظري المفترض.

H_0 (null hypothesis): The observed distribution matches the theoretically assumed distribution.

الفرض البديل (H_1): التوزيع الملاحظ لا يطابق التوزيع النظري (يختلف عنه).

H_1 (alternative hypothesis): The observed distribut-i on does not match the theoretical distribution (it differs from it).

4.1.2 Degrees of freedom 2.1.4 درجات الحرية

لإيجاد القيمة الجدولية يجب حساب درجة الحرية : (عدد الفئات- 1) بالرجوع الى جدول اختبار "كا2" الاحصائي عند مستوى الدلالة

To find the tabulated value, calculate the degrees of freedom: (number of Categories - 1). Refer to the chi-square test table at degrees of significance

Remark **ملاحظة**

من الشروط الأساسية لصلاحية تطبيق اختبار كاي-مربع لحسن المطابقة أن تكون البيانات على شكل تكرارات (عدات) مطلقة وليست نسباً مئوية أو أوساطاً حسابية.

A fundamental requirement for the validity of the chi-square test for goodness of fit is that the data be presented as absolute frequencies (counts) , not percentages or means.

كما يجب أن يكون كل تكرار متوقع (E_i) اكبر من أو يساوي 5؛ فإذا كان أقل من 5 في بعض الفئات، يمكن دمج تلك الفئات مع فئات مجاورة لرفع التكرار المتوقع. بالإضافة إلى ذلك، يتطلب الاختبار استقلالية المشاهدات، أي أن تكون كل مشاهدة مستقلة عن غيرها (مثلاً: لا يقاس نفس الفرد عدة مرات ضمن نفس الخلية). هذه الشروط تضمن دقة توزيع إحصاء كاي-مربع وصحة الاستدلال الإحصائي

Furthermore, each expected frequency (E_i) must be g reater than or equal to 5. If it is less than 5 in some categories, those categories can be combined

with adjacent categories to increase the expected frequency. Additionally, the test requires the in dependence of observations, meaning that each obse rvation must be independent of the others (e.g., t he same individual cannot be measured multiple tim es within the same cell). These conditions ensure the accuracy of the chi-square distribution and th e validity of the statistical inference.

مثال **Example**

في تجربة لعبة النرد (جميع الأوجه لها احتمال 6/1). تم رمي النرد 60 مرة، وجاءت النتائج كالتالي

In a dice game experiment (all faces have a probability of 1/6), the dice were rolled 60 times, and the results were as follows:

الوجه **face**	**1**	**2**	**3**	**4**	**5**	**6**
الملاحظ **The observer**	**7**	**13**	**8**	**15**	**6**	**11**

المطلوب: هل التوزيع الملاحظ يطابق التوزيع النظري المتساوي

Required: Does the observed distribution match the theoretically equal distribution?

الحل **Solution**

صياغة الفرضيات: **Formulating hypotheses**

الفرض الصفري:. أن احتمال ظهور كل وجه 1/6، أو بصيغة أخرى: التوزيع الملاحظ يطابق التوزيع النظري المتساوي.

Null hypothesis: probability of getting each face = 1/6, or in other words: The observed distribution matches the theoretically equal distribution.

الفرض البديل: أي أن الاحتمالات ليست متساوية لجميع الأوجه، أو أن التوزيع الملاحظ لا يطابق التوزيع النظري.

Alternative hypothesis: meaning the probabilities are not equal for all faces, or the observed distribution does not match the theoretical distribution.

اختيار الاسلوب الاحصائي المناسب : بما البيانات عبارة عن تكرارات نطبق اختار كاي مربع وحساب الصيغ التالية

Choosing the appropriate statistical method: Since the data consists of frequencies, we apply the chi-square test and calculate the following formulas.

-نحسب التكرار المتوقع تحت فرض صحة الفرضية الصفرية، يكون الاحتمال النظري لكل وجه 1/6 . عدد الرميات الكلي هو 60

إذن التكرار المتوقع لكل وجه:

We calculate the expected frequency under the assumption that the null hypothesis is true. The theoretical probability per face is 1/6. The total number of throws is 60.
Therefore, the expected frequency per face :

$$\mathbf{E_i = 60 \times (1/6) = 10}$$

All aspects $(i = 1, \dots, 6)$ جميع الأوجه.

نلخص النتائج في الجدول التالي

- We summarize the results in the following table.

الوجه (i) face	الملاحظ O_i Observed	المتوقع E_i Expected	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
1	7	10	-3	9	0.9
2	13	10	3	9	0.9
3	8	10	-2	4	0.4
4	15	10	5	25	2.5
5	6	10	-4	16	1.6
6	11	10	1	1	0.1
Sum المجموع	**60**	**60**	/		χ^2 = **6.4**

إذن: So

$$\chi^2_{\text{المحسوبة}} = 6.4$$

دراسة الدلالة المعنوية:

Significance Analysis:

<u>استخراج القيمة الجدولية:</u> لايجاد القيمة الجدولية يجب حساب درجة الحرية = (عدد الفئات- 1)= 6-1=5 بالرجوع الى جدول اختبار "كا2" الاحصائي عند درجة الحرية 5 ومستوى الدلالة **0.05** مع الاخذ بالاعتبار انها فرضية عديمة الاتجاه نجد قيمة"كا2" الجدولية**11.07** (انظر الملحق **3**)

<u>Determining the Critical Value:</u> To find the critical value, the degree of freedom (χ) must be calculated as (number of Categories - 1) = 6-1=5. Referring to the chi-square test table at **5** degrees of freedom and a significance level of **0.05,** and considering that it is a non-directional hypothesis, we find the critical chi-square value = **11.07** (See Appendix **3**)

القرار الاحصائي: بما ان قيمة"كاي2" المحسوبة اقل من القيمة الجدولية فاننا نقبل الفرض الصفري ونرفض الفرض البديل أي أن التوزيع النظري يطابق التوزيع الملاحظ (وبالتالي يطابق التوزيع المتوقع)

Statistical Decision: Since the calculated value of " χ^2" is less than the tabulated value, we accept the null hypothesis and reject the alternative hypothesis i.e., the theoretical distribution matches the observed distribution (and therefore matches the expected distribution).

2-4 اختبار كاي-2 للاستقلالية

4 -2 The Chi-Two Test for Independence

يستخدم لفحص إذا كان هناك ارتباط أو اعتماد بين متغيرين تصنيفيين (مثلاً: الجنس وتفضيل منتج، التدخين و الإصابة بمرض،... ...إلخ). بمعنى اخر هل المتغيرين مستقلين أم لا؟

يحسب بالصيغة التالية

It is used to examine whether there is a correlation or dependence between two classificatory variables (e.g., gender and product preference, smoking and disease incidence, etc.). In other words, are the two variables independent or not?

It is calculated using the following formula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where حيث

O_i:Observed frequency of class **i**	**O_i** التكرار الملاحظ للفئة **i**
E_i: Expected frequency	**E_i**: التكرار المتوقع

يحسب بالصيغة التالية :

It is calculated using the following formula:

$$\frac{\text{horizontal sum المجموع العمودي} * \text{vertical sum المجموع الافقي}}{\text{sum of total المجموع الكلي}}$$

4.1.1 Its hypotheses 1.1.4 فرضياته

الفرض الصفري (**H_0**): المتغيران مستقلان (لا علاقة بينهما او لايوجد ارتباط).

H_0 (null hypothesis): The two variables are independent (there is no relationship or correlation between them).

الفرض البديل (**H_1**): المتغيران غير مستقلين (يوجد علاقة أو ارتباط).

H_1 (alternative hypothesis): The two variables are dependent (there is a relationship or correlation between them).

4.2.2 Degrees of freedom 2.2.4 درجات الحرية

لإيجاد القيمة الجدولية يجب حساب درجة الحرية :
(عدد الاعمدة- 1) *(عدد الصفوف- 1) بالرجوع الى جدول اختبار "كا2" الاحصائي عند مستوى الدلالة

To find the tabulated value, calculate the degrees of freedom:

(number of columns - 1) *(number of rows - 1).

Refer to significace of the chi-square test table at degrees

Remark **ملاحظة**

لضمان صحة نتائج اختبار كاي-مربع **(χ^2)** يجب توفر شرطين أساسيين:

To ensure the validity of the chi-square **(χ^2)** test results two essential conditions must be met:

أولاً، أن تكون البيانات قد جُمعت من عينة عشوائية وأن تكون المشاهدات مستقلة عن بعضها البعض (أي لا يوج ارتباط بين الملاحظات).

First, the data must have been collected from a random a be independent of each o mple and the observations must ther (i.e., there must be no correlation between the observations).

ثانياً، يجب أن تكون التكرارات المتوقعة **(E_i)** مناسبة، بحيث لا تقل عن 5 في كل خلية من خلايا جدول ،أو على الأقل أن تكون 80% من الخلايا ذات تكرار متوقع اكبر او يساوي 5 وألا توجد أي خلية قيمتها المتوقعة صفر.

Second, the expected frequencies (E_i) must be appropriate with at least 5 in each cell of the table, or at least 80% of the cells having an expected frequency greater than 5, and no cell having an expected value of zero.

مثال **Example**

طرحنا سؤال على مجموعة من الطلبة في تخصصين مختلفين، هل تستخدم مراجع عربية اوفرنسية او انجليزية في دراستك

We asked a group of students in two different majors: Do you use Arabic, French, or English references in your studies

اللغة / التخصص **Majors/ Langage**	عربية **Arabic**	فرنسية **French**	الإنجليزية **English**	المجموع **Sum**
علم اجتماع **Sociology**	**25**	**20**	**10**	**55**
علم النفس **Psychology**	**10**	**18**	**30**	**58**
المجموع **Sum**	**35**	**38**	**40**	**113**

المطلوب: هل توجد فروق دالة في لغة المراجع المستخدمة من طرف طلبة التخصصين ؟

Required: Are there significant differences in the language of the references used by students in the two majors?

الحل **Solution**

صياغة الفرضيات: **Formulating hypotheses:**

الفرض الصفري: لا توجد فروق بين طلبة التخصصين في لغة المراجع المستخدمة.

Null hypothesis: There are no differences between students in the two specializations in the language of references used.

الفرض البديل: توجد فروق بين طلبة التخصصين في لغة المراجع المستخدمة.

Alternative hypothesis: There are differences between students in the two specializations in the language of references used.

اختيار الاسلوب الاحصائي المناسب : بما البيانات عبارة عن تكرارات نطبق اختار كاي مربع وحساب الصيغ التالية

Choosing the appropriate statistical method: Since the data consists of frequencies, we apply the chi-square test and calculate the following formulas

-نحسب التكرار المتوقع لكل خلية الملخص في الجدول التالي

- We calculate the expected frequency for each cell summarized in the following table

المجموع الافقي	المجموع **Sum**	الإنجليزية **English**	فرنسية **French**	عربية **Arabic**	اللغة / التخصص **Specialization / Language**
	55	**19.46/10**	**18.49/20**	**17.03/25**	**Sociology** علم اجتماع
Horizontal sum	**58**	**20.53/30**	**19.5/18**	**4.86/10**	**Psychology** علم النفس
	113	**40**	**38**	**35**	**Sum** المجموع
	الكلي **Total**		**Vertical sum**	المجموع العمودي	

التعويض في قانون Compensation in law

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{(25 - 17.03)^2}{17.03} + \frac{(10-4.86)^2}{4.86} + \frac{(20-18.49)^2}{18.49} + \frac{(18-19.5)^2}{19.5} + \frac{(10-19.46)^2}{19.46} + \frac{(30-20.53)^2}{20.53}$$

$$\chi^2 = 3.730 + 5.437 + 0.123 + 0.115 + 4.599 + 4.368$$

$$\chi^2 = 18.372$$

Significance Analysis: **دراسة الدلالة المعنوية:**

استخراج القيمة الجدولية: لايجاد القيمة الجدولية يجب حساب درجة الحرية = (عدد الاعمدة- 1)* (عدد الصفوف- 1) =4 بالرجوع الى جدول اختبار "كا2" الاحصائي عند درجة الحرية 4 ومستوى الدلالة **0.05** مع الاخذ بالاعتبار انها فرضية عديمة الاتجاه نجد قيمة"كا2" الجدولية **5.99** (انظر الملحق 3)

Determining the Critical Value: To find the critical value, the degree of freedom (χ) must be calculated as

(number of columns - 1)(number of rows - 1) = 4

Referring to the chi-square test table at **4** degrees of freedom and a significance level of **0.05**, and considering that it is a non-directional hypothesis, we find the critical chi-square value = **5.99** (See Appendix 3)

القرار الاحصائي: بما ان قيمة"كا2" المحسوبة اكبر من القيمة الجدولية فانه توجد فروق بين طلبة التخصصين في لغة المراجع المستخدمة وبالتالي نرفض الفرض الصفري ونقبل الفرض البديل

Statistical Decision: Since the calculated chi-square value is greater than the tabulated value, there are differences between students in the two specializat-

ions in the language of references used; therefore, we reject the null hypothesis and accept the alternative hypothesis.

Exercices تمارين

Exercice 1 **التمرين الاول**

يزعم عميد الكلية العلوم الاجتماعيه أن توزيع الطلاب في الأقسام الثلاثة (علم اجتماع، علم التربية، علم الحضري) هو 35 % علم اجتماع، 15 % علم التربية، 50% علم الحضري. سجلت عينة عشوائية مكونة من 200 طالب: 60 علم اجتماع، 60 علم التربية،80 علم الحضري. لمعرفة هل تتفق هذه العينة مع توزيع مدير المدرسة عند $\alpha=0.05$ ؟ اجب على الاسئلة التالية:

The Dean of the Faculty of Social Sciences claims that the student distribution across the three departments (Sociology, Education, and Urban Studies) is 35% Sociology, 15% Education, and 50% Urban Studies. A random sample of 200 students was registered: 60 Sociology, 60 Education, and 80 Urban Studies. To determine if this sample agrees with the school principal's distribution at $\alpha=0.05$, answer the following questions:

1. اكتب الفرضية الصفرية والفرضية البديلة

1. Write the null hypothesis and the alternative hypothesis.

2. احسب اختبار كاي-مربع لحسن المطابقة ماذا تستنتج

2. Calculate the chi-square test for good fit. What conclusions do you reach?

التمرين الثاني Exercice 2

نريد اجراء فرضية بحث تتطلب الكشف عما إذا كان هناك استقلالية اختيار التخصص الدراسي عن الجنس لدى طلبة شعبة علم الاجتماع ، تحصل الباحث على البيانات التالية حسب الجدول الآتي

We want to conduct a research hypothesis that requires investigating whether there is an-independence between the choice of academic specialization and gender among sociology students. The researcher obtained the following data according to the table below.

التخصص/الجنس Specialization/Gender	علم التربية Educational science	علم الحضري Urban science	تنظيم وعمل Organization and operation
اناث	30	65	120
ذكور	50	75	55

فسر الفرضية الصفرية عند $\alpha = 0.010$

Explain the null hypothesis when $\alpha = 0.010$

Chapter 5 الفصل الخامس

نماذج الانحدار الخطي

Linear Regession models

تمهيد Preface

الانحدار الخطي يعتبر واحدة من الأدوات الإحصائية الأساسية المستخدمة لنمذجة العلاقة بين متغير تابع و متغير مستقل أو أكثر بطريقة خطية، بهدف تحليل تلك العلاقات والتنبؤ بالقيم المستقبلية. يستند هذا النموذج إلى فكرة أن الظواهر الطبيعية والإنسانية ليست عشوائية بشكل كامل، بل تتبع أنماطاً يمكن الكشف عنها وتحليلها. يتميز ببنية رياضية بسيطة وتفسير واضح، حيث يُمثَّل بخط مستقيم في أبسط حالاته **(الانحدار البسيط)** أو بمستوى متعدد الأبعاد في السيناريوهات الأكثر تعقيدا **(الانحدار المتعدد)**. تتعدد استخداماته ليشمل مجالات مثل الاقتصاد، الطب، العلوم الاجتماعية، والهندسة، مايجعله أداة فعالة تربط البيانات المجردة بالقرارات الواقعية. بالرغم من أنه يوضح العلاقة بين المتغيرات دون إثبات السببية بشكل قطعي، يظل أساساً مهماً في التحليل الإحصائي والنماذج التنبؤية، جامعاً بين الدقة الرياضية وسهولة التطبيق مع القدرة على مواجهة تحديات البيانات الحديثة

Linear regression is a fundamental statistical tool used to model the relationship between a dependent variable and one or more independent variables in a linear fashion, with the aim of analyzing these relationships and predicting future values. This model is based on the idea that natural and human phenomena are not entirely random, but rather follow patterns that can be detected and analyzed. It is characterized by a simple mathematical structure and clear interpretation, being represented by a straight line in its simplest form **(simple regression)**or by a multidimensional plane in more complex scenario (**multiple regression**) . Its applications are diverse, encompassing fields such as economics, medicine, social sciences, and engineering, making it an effective tool for connecting abstract data to real-world decisions. Although it illustrates the relationship between variables without definitively proving causation, it remains a crucial foundation in statistical analysis and predictive models, combining

mathematical precision and ease of application with the ability to address the challenges of modern data.

5-1 اساسيات الانحدار الخطي

5-1 Fundamentals of Linear Regression

Definition of regression **تعريف الانحدار**

هو أسلوب يمكن بواسطته تقدير قيمة أحد المتغير بمعلومية قيمة المتغير الآخرعن طريق **معادلة الانحدار**

It is a method by which the value of one variable can be estimated given the value of the other variable using **regression equation**

Type of regression **أنواع الانحدار**

الانحدار الخطي البسيط :فكلمة " **بسيط** " تعني أن المتغير التابع **Y** يعتمد على متغير مستقل واحد وهو**X** وكلمة " **خطي** " تعني أن العلاقة بين المتغيرين(X , Y) **علاقة خطية** و تكتب معادلته على النحو التالي

Simple linear regression: The word **"simple"** means that the dependent variable Y depends on only one independent variable, **X**, and the word **"linear"** means that the relationship between the two variables (**Y** and **X**) is linear. Its equation is written as follows:

$$\mathbf{i=1,..,n} \quad / \quad Y_i = \alpha + \beta X_i + \varepsilon_i$$

where **حيث**

α و β معلما النموذج أما ε_i فيمثل الخطأ في تفسير Y_i

α and β are parameters of the model, while ε_i represents the error in interpreting Y_i.

الانحدار الخطي المتعدد : إذا كان المتغير **Y** يعتمد على أكثر من متغير مستقل و تكتب معادلته على النحو التالي

Multiple linear regression: This occurs when the variable **Y** depends on more than one independent variable. Its equation is written as follows:

$$\mathbf{i=1,..,n} \quad / \quad \mathbf{Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_2 + \varepsilon_i}$$

الانحدار غير الخطي : إذا كانت العلاقة بين المتغير **Y** والمتغيرات المستقلة غير خطية كأن تكون من الدرجة الثانية أو آسية ،مثل كالصيغة التالية

Nonlinear regression: This occurs when the relationship between the variable Y and the independent variables is nonlinear, such as being of the second degree or exponential.

For example, like the following formula

$$\mathbf{i = 1,..,n \;/\; Y_i = \alpha + \beta_1 X_1^2 + \beta_2 X_2^2 + \ldots + \beta_i X_i^2 + \varepsilon_i}$$

خطوات بناء نموذج الانحدار

Steps for building a regression model

-جمع البيانات واستكشافها
- Data collection and exploration

-التحقق من الافتراضات : كاختبار الخطية
- Testing assumptions: such as linearity testing

-تقدير النموذج : حساب معاملات الانحدار باستخدام طريقة المربعات الصغرى و تفسير المعاملات

- Model estimation: calculating regression coefficients using the least squares method and interpreting the coefficient

-تقييم النموذج : و ذلك بحساب معامل التحديد R^2

- Model evaluation: calculating the coefficient of determination **(R^2)**

- Prediction and inference -التنبؤ والاستنتاج

5-2 تحليل الانحدار الخطي البسيط

5 -2 Simple Linear Regression analysis

في عالم يزداد تعقيداً، يظل الانحدار الخطي البسيط أداة قوية لأنه بسيط في الفهم والتطبيق واضح في النتائج والتفسير فهو بوصلة الإحصائي الأولى في رحلة استكشاف عالم البيانات المعقد سنخصص في هذا الجزء مراحل تحليل الانحدار الخطي البسيط

In an increasingly complex world, simple linear regression remains a powerful tool because it is easy to understand and apply, and its results and interpretation are clear. It is the

statistician's primary compass in the journey of exploring the complex world of data. In this section, we will focus on the stages of simple linear regression analysis

5-2-1 تقدير معاملات نموذج الانحدار الخطي البسيط

5-2-1 Estimating the parameters of a simple linear regression model

بعد تمثيل الأزواج المرتبة على مستوي متعامد ومتجانس نحصل على شكل الانتشار فإذا أظهر الشكل الانتشاري للبيانات أن هناك علاقة خطية بين المتغيرين نقوم بتقدير خط الانحدار **Y**على **X** بواسطة العلاقة التالية

After representing the ordered pairs on a homogeneous plane, we obtain the scatter plot. If the scatter plot of the data shows a linear relationship between the two variables, we estimate the regression line Y/X using the following relationship

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

تحسب القيمتان $\hat{\alpha}$ و $\hat{\beta}$ باستخدام طريقة المربعات الصغرى حيث

The two valuese $\hat{\alpha}$ **and** $\hat{\beta}$ are calculated using the least squar method where

$$\widehat{\beta} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \quad , \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X}$$

يمكن بعد ذلك إجراء تقييم شامل لاختبار الفرضيات المتعلقة بالمعلمات، كما سيتم توضيحه في المثال المقبل

A comprehensive evaluation of the hypothesis testing related to the parameters can then be carried out, as will be illustrated in the next example

Example **مثال**

لدراسة العلاقة بين ساعات الدراسة والدرجات المتحصل عليها لمجموعة من الطلبة ،كانت النتائج كالتالي

To study the relationship between study hours and grades obtained for a group of students, the results were as follow

hours الساعات (X)	32	41	31	35	26	21
Grade(y)الدرجات	6	12	5	6	4	3

احسب تقديرات معادلة الإنحدار ومن ثم إختبر الفرضيات عن معنويات المعلمات

Calculate the regression equation estimates and then test the hypotheses about the significance of the parameters

Solution **الحل**

1-إنشاء جدول الحسابات

1-Create the required chart

i	X	Y	X^2	Y^2	XY
1	32	6	1024	36	192
2	41	12	1681	144	492
3	31	5	961	25	155
4	35	6	1225	36	210
5	26	4	676	16	104
6	21	3	441	9	63
المجموع Sum	$\Sigma X = 186$	$\Sigma Y = 36$	$\Sigma X^2 = 6008$	$\Sigma Y^2 = 266$	$\Sigma XY = 1216$

بناءا على ذلك فإن المعلمات المقدرة

Therefore, the estimated parameters

$$\hat{\beta} = \frac{6(1216)-(186)(36)}{6(6008)-(186)^2} = \frac{600}{1452}$$

$$\hat{\beta} \approx 0.4132$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\,\bar{X} = \frac{36}{6} - 0.4132\left(\frac{186}{6}\right)$$

$$\hat{\alpha} = \frac{-40.8552}{6}$$

$$\hat{\alpha} \approx -6.809$$

المعادلة المقدرة Estimated equation

$$\hat{Y}_i = -6.809 + 0.4132\,X_i$$

المعادلة تعني ان كل زيادة بمقدار ساعة واحدة في الدراسة تؤدي الى زيادة متوقعة بمقدار **0.4312**

كما ان قيمة المتغير التابع**6.809-** إذا كان المتغير المستقل مساويا للصفر

The equation means that every one-hour increase in study time leads to an expected increase of **0.4312** degree. Also, The value of the dependent variable is equal **-6.809** if the independent variable is equal to zero.

2-تقييم معادلة الإنحدار: إختبار الفرضيات لمعنويات المعلمات ومعامل التحديد

2-Evaluating the regression equation: Hypothesis testing for parameter significance and coefficient of determination.

بعد الحصول على معادلة الإنحدار المقدرة يجب تقييمها

بإجراء إختبار الفرضيات لمعنويات كل من α وβ

و الحصول على معامل التحديد R^2

After obtaining the estimated regression equation, it must be evaluated by conducting hypothesis testing for the significance of both α and β obtain the coefficient of determination R^2

- **إختبار الفرضية لمعنوية المعامل α و β**

-Testing the hypothesis for coefficient β and α significance

Formulating the hypothesis **صياغة الفرضية**

$H_0 : \alpha = 0$	$H_0 : \beta = 0$
$H_1 : \alpha \neq 0$	$H_1 : \beta \neq 0$

تحديد منطقتي القبول والرفض بتحديد درجة الحرية **df=n-2** وإستخراج قيمة الجدولية عند درجة الأهمية **5 %** وهي **2.776**

The acceptance and rejection zones are determined by setting the degree of freedom **df=N-2** and extracting the tabulated value at a significance level of **5% ,** which is **2.776.**

تحديد القيمة المحسوبة t

Determine the calculated value t

تحسب بالقانون التالي

It is calculated according to the following law

$$\mathbf{t_2 = \frac{\hat{\beta}}{\sigma_\beta}} \quad , \mathbf{t_1 = \frac{\hat{\alpha}}{\sigma_\alpha}}$$

حيث $\boldsymbol{\sigma_\alpha}$ و $\boldsymbol{\sigma_\beta}$ الخطأ المعياري للمعلمات α و β

Where $\boldsymbol{\sigma_\beta}$and $\boldsymbol{\sigma_\alpha}$ the standard error of the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

يمكن إستخراج قيمة الخطأ المعياري لكل من α و β على النحو التالي بإستخدام الجدول التالي والمعادلات التالية

The standard error values for both α and β can be extracted using the following table and equations

$$\sigma_{\beta}=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{(n-2)\sum(X_i-\bar{x})^2}} \quad , \quad \sigma_{\alpha}=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{(n-2)}\frac{\sum X^2}{n\sum(x_i-\bar{x})^2}}$$

i	X_i	Y_i	$\hat{Y}_i$	$(X_i - \bar{x})$	$(X_i-\bar{x})^2$	$(Y_i-\hat{Y}_i)^2$
1	32	6	6.41	1	1	0.1681
2	41	12	10.13	10	100	3.4969
3	31	5	5.99	0	0	0.9801
4	35	6	7.65	4	16	2.7225
5	26	4	3.93	-5	25	0.0049
6	21	3	1.87	-10	100	1.2769
المجموع Sum	**186**	**36**	-	**0**	**242**	**8.6494**

And from it **ومنه**

$$\sigma_{\beta} = \sqrt{\frac{8.6494}{(6-2)(242)}} = \sqrt{\frac{8.6494}{968}} \approx 0.0945$$

$$\sigma_{\alpha} = \sqrt{\frac{8.6494}{(6-2)}\frac{6008}{6(242)}} = \sqrt{(2.1623)(4.1377)} \approx 2.9911$$

$$t_1 = \frac{\hat{\alpha}}{\sigma_{\alpha}} = \frac{-6.809}{2.9911} \approx -2.2764$$

$$t_2 = \frac{\hat{\beta}}{\sigma_{\beta}} = \frac{0.4132}{0.0945} \approx 4.3724$$

We conclude the following **نستنتج مايلي**

$\|t_1\| = 4.372 > 2.776$ H_0 نرفض (Refuse H_0)	$\Rightarrow$ معنوي إحصائيا β

	(statistical significance)
$\|t_2\|$ =2.2764 <2.776 H_0 نقبل (Accept H_0)	⇒غير معنوي احصائيا α (Non Statistical singficance)

3-معامل التحديد R^2

3-Coefficient of determination R^2

معامل التحديد يشرح جودة المعادلة و يظهر النسبة المفسر ة من التغيرات في المتغير التابع . تتراوح قمتهما بين الواحد الصحيح والصفر أي أن $0 \leq R^2 \leq 1$ ويمكن حسابه بالمعادلة التالية :

The regression coefficient explains the quality of the equation; it shows the proportion of changes in the dependent variable that are explained 0 and 1

i.e $0 \leq R^2. \leq 1$

It can be calculated using the following equation

$$R^2 = \frac{\alpha \sum Y + \beta \sum XY - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

$$= \frac{-6.809(36) + 0.4132(1216) - 6(6)^2}{266 - 6(6)^2}$$

$$R^2 \approx \frac{41.3272}{50} \approx 0.8265$$

Explanation of R^2 **تفسير R^2**

82.6% من التباين في الدرجات يمكن تفسيره بساعات الدراسة اي أن النموذج جيد نسبياً في تفسير العلاقة

of the variance in scores can be explained by **82.6**% study hours, meaning

model is relatively good at explaining the the relationship

4-حساب التوقعات والتنبؤات

4 - Forecast and prediction calculation

لحساب التوقع نقوم بتقدير قيمة المتغير التابع Y_f عندما تُعطى القيمة الفعلية أو المُخططة للمتغير المستقل X_f ويمكن أن تكون معادلة الإنحدار المقدرة بالنسبة للمخطط والمتوقع كالتالي :

To calculate the expected value, we estimate the value of the dependent variable Y_f when given the actual or planned value of the independent variable X_f. The estimated regression equation for the planned and expected values can be as follows:

$$\hat{Y}_f = \hat{\alpha} + \hat{\beta} X_f$$

مثلا حساب كم يحصل طالب يدرس 30 ساعة من درجة **أي** $X_f = 30$

For example, calculating how many grades a student who studies 30 hours will receive

i.e, $X_f = 30$

$$\hat{Y} = -6.809 + 0.4132 \times 30$$

$$= -6.809 + 12.396$$

$$5.587 \approx 5.6$$

اي يحصل على **5.6** درجة

He receives a grade of **5.6**

5-2-2 تقدير معاملات نموذج الانحدار الخطي المتعدد

5.2. 2 Estimating the coefficients of a multiple linear regression model

في الانحدار المتعدد، يُعبّر التوقع الشرطي عن متوسط المتغير التابع Y عند قيم ثابتة للمتغيرات المستقلة X_1 و X_2، حيث يمثل الثابت α متوسط أثر المتغيرات غير المدرجة في النموذج .أما معامل الميل β_1 فيقيس التغيرفي متوسط Yلكل وحدة تغير في X_1 مع تثبيت X_2، أي التأثير المباشر أو الصافي لـ X_1، وبالمثل β_2 يقيس التأثير المباشر لـ X_2 مع تثبيت X_1.

In multiple regression, the conditional expectancy represents the mean of the dependent variable **Y** at constant values for the independent variables X_1 and X_2, where the constant **α** represents the mean effect of the variables not included in the model.

The slope coefficient β_1 measures the change in mean **Y** per unit change in X_1 with X_2 held constant, i.e., the direct or net effect of X_1.

Similarly, β_2 measures the direct effect of X_2 with X_1 held constant.

يمكن تقدير معادلة الإنحدار ذات الثلاث متغيرات على النحو التالي

The three-variable regression equation can be estimated as follows:

$$\widehat{\beta_1} = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2)}$$

$$\widehat{\beta_2} = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\widehat{\alpha} = \bar{Y} - \widehat{\boldsymbol{\beta_1}}\ \bar{X}_1 - \widehat{\boldsymbol{\beta_2}}\ \bar{X}_2$$

مثال **Example**

الجدول يوضح درجة الطالب النهائية **Y** ،عدد ساعات الدراسة X_1 ودرجة الذكاء X_2 لخمسة من الطلاب

The table shows the final grade **Y**, number of study hours X_1, and scores of intelligence X_2 of five students.

i	Y	X_1	X_2
1	3	2	1
2	5	3	2
3	6	4	2
4	7	5	3
5	9	6	4

المطلوب تقدير دالة الإنحدار التي تفسر العلاقة ما بين درجة الطالب النهائية وأثر كل من نسبة عدد ساعات الدراسة ودرجة الذكاء .تقديم التحليل اللازم والتقييم للنتائج.

The task is to estimate the regression function that explains the relationship between the student's final grade and the effect of both the number of study hours and score of intelligence. Then, provide the necessary analysis and evaluation of the results.

الحل **Solution**

1-إنشاء جدول الحسابات **1-Create the required chart**

i	Y	X_1	X_2	X_1 X_2	X_2^2	Y X_1	Y X_2
1	1	2	3	6	9	2	3
2	5	3	2	6	4	15	10
3	6	4	2	8	4	24	12
4	7	5	3	15	9	35	21
5	9	6	4	24	16	54	36
المجموع Sum	30	20	12	55	34	134	82

2-تقدير معادلة الانحدار

2- Estimating the regression equation

يمكن تلخيص ذلك كما يلي :

This can be summarized as follows:

$$\widehat{\boldsymbol{\beta_1}} = \frac{(134 \times 34) - (82 \times 55)}{(90 \times 34) - (55)^2} = \frac{\mathbf{46}}{\mathbf{35}} \approx \mathbf{1.32}$$

$$\widehat{\boldsymbol{\beta_2}} = \frac{(82 \times 90) - (134 \times 55)}{(90 \times 34) - (55)^2} = \frac{\mathbf{10}}{\mathbf{35}} \approx \mathbf{0.29}$$

$$\hat{\alpha} = \bar{Y} - \widehat{\boldsymbol{\beta_1}}\ \bar{X}_1 - \widehat{\boldsymbol{\beta_2}}\ \bar{X}_2$$

$$\hat{\alpha} = 6 - (1.32 \times 4) - (0.29 \times 2.4)$$

$$= 6 - 5.26 - 0.69$$

$$\mathbf{= 0.0571 \approx 0.057}$$

اذن **معادلة الانحدار**

Therefore, the **regression equation**

$$\hat{Y} = 0.057 + 1.32\ X_1 + 0.29\ X_2$$

تفسير معادلة الانحدار

Interpretation of the regression equation

$\alpha = 0.057$: عندما يدرس الطالب صفر ساعة وكانت درجة ذكائه صفراً (وهو أمر افتراضي غير واقعي)، فإن درجته المتوقعة تكون **0.057** (تقريباً صفر). هذا يدل على أنه بدون دراسة وبدون ذكاء، لا يُتوقع الحصول على درجة تذكر.

$\alpha = 0.057$ when a student studies for zero hours and their intelligence score is zero (a hypothetical and unrealistic assumption), their expected score is **0.057** (approximately zero). This indicates that without studying and without intelligence, no significant score can be expected.

$\beta_1 = 1.32$ (معامل ساعات الدراسة):

بافتراض ثبات درجة الذكاء، كل زيادة بمقدار ساعة واحدة في الدراسة تؤدي إلى زيادة في الدرجة النهائية بمقدار **1.32** درجة. هذا يظهر أن للدراسة تأثيراً إيجابياً قوياً نسبياً على التحصيل.

$\beta_1 = 1.32$ (Study Hours Coefficient):

Assuming constant intelligence, each additional hour of study results in a **1.32** point increase in the final score. This demonstrates that studying has a relatively strong positive effect on achievement.

$\beta_2 = 0.29$ (معامل درجة الذكاء):

بافتراض ثبات ساعات الدراسة، كل زيادة بمقدار درجة واحدة في مقياس الذكاء (وحدة واحدة من X_2) تؤدي إلى زيادة في الدرجة النهائية بمقدار **0.29** درجة. التأثير هنا أضعف من تأثير ساعات الدراسة (لكل وحدة).

β_2 = 0.29 (Intelligence Score Coefficient):

Assuming constant study hours, each one-point increase in the intelligence score (one X_2 unit) results in a **0.29** point increase in the final score. The effect here is weaker than the effect of study hours (per unit).

تمارين Exercices

التمرين الأول Exercice 1

لدراسة تاثير الاستهلاك المحلي **(y)** على الانتاجية **(x)** على مدى عدة سنوات، أخذنا عشر قراءات تقريبية على النحو التالي:

To study the impact of domestic consumption **(y)** and productivity **(x)** over several years, we took ten approximate readings as follows:

X	**12**	**10**	**8**	**18**	**29**	**14**	**20**	**33**	**19**	**32**
Y	**7**	**5**	**6**	**9**	**8**	**11**	**14**	**10**	**6**	**4**

1 . اكتب معادلة الانحدار

1. Write the regression equation.

2. استخدم المعادلة للتنبؤ بقيمة الانتاج عندما يكون قيمة الاستهلاك 38

2.Use the equation to predict the value of production when the value of consumption is **38**

3. احسب معامل التحديد R^2 وفسره

3. Calculate and interpret the coefficient of determination (R^2).

التمرين الثاني Exercice2

ارادت مؤسسة للإعلان تأثير عدد المستجيبين للإعلان X_1 و حجم الاعلان المنشور X_2 على عدد الصحف التي تنشر الاعلان **Y** وحصلت على النتائج التالية

An advertising agency wanted to determine the effect of the number of respondents to an advertisement (X_1) and the size of the advertisement (X_2) on the number of newspapers that publish the advertisement **(Y)**, and obtained the following results

Y	**1**	**8**	**3**	**5**	**6**	**10**	**25**	**16**
X_1	**2**	**8**	**1**	**7**	**4**	**6**	**7**	**9**
X_2	**1**	**4**	**1**	**3**	**2**	**4**	**5**	**8**

1- احسب دالة الإنحدار التي تفسر العلاقة ما بين عدد الصحف وأثر كل من عدد المستجيبين للإعلان وحجم الاعلان المنشور .

1- Calculate the regression function that explains the relationship between the number of newspapers and the effect of both the number of respondents to the advertisement and the size of the published advertisement.

2- قدم التحليل اللازم والتقييم للنتائج.

2- Provide the necessary analysis and evaluation of the results.

التمرين الثالث Exercice3

البيانات التالية أخذت عينة من 7 أيام من مصنع لدراسة تأثير درجة الحرارة X_1 وضغط التشغيل X_2 على الإنتاجية **Y** :

The following data were taken from a 7-day sample taken from a factory to study the effect of temperature (X_1) and operating pressure (X_2) on productivity **(Y)**. The sample consisted of 7 days:

Day اليوم	درجة الحرارة temperature (X_1)	ضغط التشغيل operating pressure (X_2)	الإنتاجية productivity (Y)
1	**36**	**78**	**12**
2	**25**	**91**	**95**
3	**35**	**65**	**80**
4	**31**	**74**	**72**
5	**40**	**69**	**82**
6	**38**	**32**	**92**
7	**41**	**71**	**65**

1- احسب معادلة الانحدار الخطي البسيط لـ **Y** على X_1 وحدها، ثم لـ **Y** على X_2 وحدها.

1- Calculate the simple linear regression equation for Y on X_1 alone, and then for Y on X_2 alone.

2- احسب معادلة الانحدار المتعدد لـ **Y** على X_1 و X_2 معاً.

2 -Calculate the multiple regression equation for Y on both X_1 and X_2 together.

3- ما هي الإنتاجية المتوقعة عند درجة حرارة 70 وضغط 10 ؟ (باستخدام نموذج الانحدار المتعدد).

3- What is the expected yield at a temperature of 70°C and a pressure of 10? (Using the multiple regression model)

ملاحق

Appendices

الملحق (1): جدول التـــــوزيـــــــــــع الطبيعي المعياري

Appendix (1): Standard Normal Distribution Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

الملحق (2): جدول التـــــوزيـــــــــع ستـــــودنـــــت

Appendix (2): Student Distribution Table

df	$t_{0.10}$	$t_{0.50}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

Appendix 3: Chi-square Distribution Table

df	α 0.995	α 0.990	α 0.975	α 0.95	α 0.90	α 0.1	α 0.05	α 0.025	α 0.010	α 0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Appendix 4: Pearson Distribution Tabl

Level of Significance for a One-Tailed Test مستوى الدلالة لاختبار ذي طرف واحد

0.10	0.05	0.02	0.01	0.005	0.0005

Level of Significance for a Two-Tailed Test مستوى الدلالة لاختبار ذي طرفين

df	0.20	0.10	0.05	0.02	0.01	0.001
1	0.951	0.988	0.997	0.9995	0.9999	0.99999
2	0.800	0.900	0.950	0.980	0.990	0.999
3	0.687	0.805	0.878	0.934	0.959	0.991
4	0.608	0.729	0.811	0.882	0.917	0.974
5	0.551	0.669	0.755	0.833	0.875	0.951
6	0.507	0.621	0.707	0.789	0.834	0.925
7	0.472	0.582	0.666	0.750	0.798	0.898
8	0.443	0.549	0.632	0.715	0.765	0.872
9	0.419	0.521	0.602	0.685	0.735	0.847
10	0.398	0.497	0.576	0.658	0.708	0.823
11	0.380	0.476	0.553	0.634	0.684	0.801
12	0.365	0.457	0.532	0.612	0.661	0.780
13	0.351	0.441	0.514	0.592	0.641	0.760
14	0.338	0.426	0.497	0.574	0.623	0.742
15	0.327	0.412	0.482	0.558	0.606	0.725
16	0.317	0.400	0.468	0.542	0.590	0.708
17	0.308	0.389	0.456	0.529	0.575	0.693
18	0.299	0.378	0.444	0.515	0.561	0.679
19	0.291	0.369	0.433	0.503	0.549	0.665
20	0.284	0.360	0.423	0.492	0.537	0.652
21	0.277	0.352	0.413	0.482	0.526	0.640
22	0.271	0.344	0.404	0.472	0.515	0.629
23	0.265	0.337	0.396	0.462	0.505	0.618
24	0.260	0.330	0.388	0.453	0.496	0.607
25	0.255	0.323	0.381	0.445	0.487	0.597
26	0.250	0.317	0.374	0.437	0.479	0.588
27	0.245	0.311	0.367	0.430	0.471	0.579
28	0.241	0.306	0.361	0.423	0.463	0.570
29	0.237	0.301	0.355	0.416	0.456	0.562
30	0.233	0.296	0.349	0.409	0.449	0.554
40	0.202	0.257	0.304	0.358	0.393	0.490
60	0.165	0.211	0.250	0.295	0.325	0.408
120	0.117	0.150	0.178	0.210	0.232	0.294
∞	0.057	0.073	0.087	0.103	0.114	0.146

ملحق 5 : جـــــدول توزيـــــــــــع سبيـــــرمـــــان

Appendix 5: Spearman Distribution Table

	α				
n	0.10	0.05	0.025	0.01	0.005
5	0.800	0.900	1.000	1.000	-
6	0.657	0.829	0.886	0.943	1.000
7	0.571	0.714	0.786	0.893	0.929
8	0.524	0.643	0.738	0.833	0.881
9	0.483	0.600	0.700	0.783	0.833
10	0.455	0.564	0.648	0.745	0.794
11	0.427	0.536	0.618	0.709	0.755
12	0.406	0.503	0.587	0.678	0.727
13	0.385	0.484	0.560	0.648	0.703
14	0.367	0.464	0.538	0.626	0.679
15	0.354	0.446	0.521	0.604	0.654
16	0.341	0.429	0.503	0.582	0.635
17	0.328	0.414	0.488	0.566	0.618
18	0.317	0.401	0.472	0.550	0.600
19	0.309	0.391	0.460	0.535	0.584
20	0.299	0.380	0.447	0.522	0.570
21	0.292	0.370	0.436	0.509	0.556
22	0.284	0.361	0.425	0.497	0.544
23	0.278	0.353	0.416	0.486	0.532
24	0.271	0.344	0.407	0.476	0.521
25	0.265	0.337	0.398	0.466	0.511
26	0.259	0.331	0.390	0.457	0.501
27	0.255	0.324	0.383	0.449	0.492
28	0.250	0.318	0.375	0.441	0.483
29	0.245	0.321	0.368	0.433	0.475
30	0.240	0.306	0.362	0.425	0.467

خاتمة Conclusion

بعد استعراضنا لمفاهيم هذا المطبوعة، من الأسس النظرية إلى التطبيقات العملية، يمكننا القول إن الإحصاء الاستدلالي ليس مجرد مجموعة من الصيغ الرياضية أو الاختبارات الجافة، بل هو أداة فكرية ومنهجية تمكن الباحث من تجاوز حدود بيانات العينة للوصول إلى استنتاجات معممة حول مجتمعات أكبر. إنه الجسر الذي يربط الملاحظة الجزئية بالحقيقة الكلية، ويمنح البحث العلمي شرعيته وقوته الاستدلالية.

After reviewing the concepts in this publication, from theoretical foundations to practical applications, we can say that inferential statistics is not merely a set of mathematical formulas or dry tests. Rather, it is an intellectual and methodological tool that enables researchers to transcend the limitations of sample data to reach generalizable conclusions about larger populations. It is the bridge that connects partial observation to overall truth, and it grants scientific research its legitimacy and inferential power.

كبداية بدأنا بمفاهيم أساسية للاحصاء الاستدلالي والمبادئ الفلسفية لاختبار الفرضيات. وأوضحنا أن كل اختبار إحصائي يقوم على افتراضات لا بد من التحقق منها، وأن اختيار الاختبار المناسب يظل القرار الأكثر حسما في أي دراسة تجريبية.

begin, we started with the basic concepts of inferential statistics and the philosophical principles of hypothesis testing. We explained that every statistical test is based on assumptions that must be verified, and that choosing the appropriate test remains the most crucial decision in any experimental study.

ثم انتقلنا إلى التطبيقات العملية: الجداول التكرارية بأنواعها، معاملات الارتباط (بيرسون، سبيرمان، كيندال، والجزئي)، اختبارات كاي-مربع (لحسن المطابقة

والاستقلالية)، ونماذج الانحدار الخطي البسيط والمتعدد. وتوقفنا عند كل أداة لنبين متى تستخدم، وكيف تحسب نتائجها، والأهم كيف تفسر هذه النتائج في سياق البحث.

then moved on to practical applications: frequency tables of various types, correlation coefficients (Pearson, Spearman, Kendall, and partial), chi-square tests (for goodness of fit and independence), and simple and multiple linear regression models. We examined each tool to show when to use it, how to calculate its results, and, most importantly, how to interpret these results within the context of the research.

في النهاية، هذه المطبوعة ليس نهاية الطريق في تعلم الإحصاء، بل بوابة نحو تفكير إحصائي نقدي. الباحث المتمكن لا يحفظ الاختبارات، بل يعرف متى يستخدم كل اختبار، ولماذا يختاره، وكيف يفسر مخرجاته، وما هي حدود ما يمكنه قوله بناءً عليها. ومع التطور المتسارع في أدوات تحليل البيانات ، تظل المبادئ التي تعلمناها هنا – دقة القياس، فهم الافتراضات، والتقديرات هي الركيزة الثابتة لكل استدلال علمي سليم.

Ultimately, this publication is not the end of the road in learning statistics, but rather a gateway to critical statistical thinking. A competent researcher doesn't simply memorize tests; they know when to use each test, why they choose it, how to interpret its results, and the limits of what they can conclude. And despite the rapid advancements in data analysis tools, the principles we've learned here measurement accuracy, understanding assumptions, and estimations remain the bedrock of all sound scientific reasoning.

ختامًا، نشجع طالب على التعامل مع الإحصاء الاستدلالي كأداة للاكتشاف لا للتعقيد. إن فهم هذا المطبوعة وتطبيقها بوعي سيساهم في بناء علم أكثر شفافية وموثوقية. وفقكم الله في مسيرتكم البحثية.

In conclusion, we encourage every student to approach inferential statistics as a tool for discovery, not complexity. Understanding and consciously applying this publication will contribute to building a more transparent and reliable science. May God grant you success in your research endeavor

References مـر اجع

Arabic references المراجع العربية

1. إبراهيم علي إبراهيم عبد ربه، مبادئ علم الإحصاء، الدار الجامعية، مصر، 2002
2. اياد محمد الهوبي، الاحصاء التطبيقي، الكلية الجامعة للعلوم والتكنولوجيا، الطبعة الاولى، 2012.
3. حسن ياسين طعمة وإيمان حسين حنوش. الإحصاء الاستدلالي. الطبعة الأولى، دار صفاء للنشروالتوزيع: عمان، 2012
4. جمال رشيد الكحلوت. مبادئ نظرية العينات. مكتبة الفريد الالكترونية، 2003
5. علي أحمد السقاف. الاحصاء الوصفي والاستدلالي. الطبعة الأولى، إصدار المركز الديمقراطي العربي، برلين المانيا،
6. شبيجل وآخرون. الاحتمالات والإحصاء: ملخصات شوم إيزي. الدار الدولية للاستثمارات الثقافية،مصر، د ت ن ، ص.26
7. **بوحفص عبد الكريم: الإحصاء المطبق في العلوم الاجتماعية والإنسانية، ديوان المطبوعات الجامعية، الجزائر، 2005.**
8. **حسن محمد حسن: مبادئ الإحصاء الاجتماعي، دار المعرفة الجامعية، 2000.**
9. **أحمد عبد السميع طيبة: مبادئ الإحصاء (الوصفي و الاستدلالي):**
   https://drive.google.com/.../0B7LWyrvVvOnuYjNPTVZ3VG.../view
10. **إياد محمد الهوبي: الإحصاء التطبيقي:**

    https://drive.google.com/.../0B7LWyrvVvOnuVnpQWlc1MV.../view .

Foreign references المراجع الاجنبية

1. Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the Behavioral Sciences*. Cengage Learning.
2. Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2023). *Probability and Statistical Inference* (11th ed.). Pearson. -
3. Howell, D. C. (2017). *Fundamental Statistics for the Behavioral Sciences*. Cengage Learning.
4. Lehmann, E. L., & Romano, J. P. (2022). *Testing Statistical Hypotheses* (4th ed.). Springer.
5. Hinton, P. R. (2025). *Statistics Explained* (4th ed.). Routledge
6. Longford, N. T. (2021). *Statistics for Making Decisions*. CRC Press.
7. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). John Wiley & Sons.
8. Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). Sage Publications.
9. Bonett, D. G., & Wright, T. A. (2000). "Sample size requirements for estimating Pearson, Kendall and Spearman correlations". *Psychometrika*, 65, 23-28.
10. Casella, G., & Berger, R. L. (2021). *Statistical Inference* (2nd ed.). Cengage Learning.